THE MECHANISTIC INVESTIGATION OF THE BACTERIAL THIAMIN THIAZOLE SYNTHASE AND THE EUKARYOTIC THIAMIN PYRIMIDINE SYNTHASE

A Dissertation

by

RUNG-YI LAI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Tadhg P. Begley
Committee Members,	Frank M. Raushel
	Wenshe Liu
	Gregory D. Reinhart
Head of Department,	Francois P. Gabbai

August 2015

Major Subject: Chemistry

ABSTRACT

Thiamin pyrophosphate is an essential cofactor in all living systems. Its biosynthesis involves two separate pathways to synthesize pyrimidine and thiazole, followed by a coupling of these two heterocycles to generate thiamin. In this dissertation, the *Bacillus subtilis* thiazole synthase (ThiG) and *Candida albicans* pyrimidine synthase (THI5p) were investigated.

The unanticipated reverse reaction of ThiG and the thiazole tautomer generated an intermediate, which was characterized by mass spectrometry and further applied to investigate the later steps of thiazole biosynthesis involving ThiG and ThiO, which is the glycine oxidase. This study demonstrated our proposed mechanism of *B. subtilis* thiazole biosynthesis involving ThiG and ThiO.

The THI5 protein (THI5p) was reconstituted *in vitro* and identified as a single turnover protein to use an active-site histidine to react with pyridoxal 5'-phosphate (PLP) to generate HMP-P. The histidine structure after HMP-P formation was characterized as a keto-acid. The PLP by-products were trapped by phenylhydrazine derivatization. The structural characterization identified as glyoxylate via C3-fragment. The keto-acid, C3-fragment and glyoxylate had oxygen incorporations from O_2. The pattern of incorporations suggested that the protein utilized multiple O_2 to synthesize HMP-P. Moreover, a shunt product, PLP-nitrile, was identified to suggest that the reaction was initiated by a radical reaction. Based on these preliminary information, a mechanism was proposed. To investigate the proposed mechanism, phosphopeptide

enrichment was used to identify the peptides containing possible shunt intermediates. The investigations could help to test the proposed mechanism.

Because THI5p required Fe(II) and O_2 to catalyze HMP-P synthesis, preliminary bioinorganic investigations were conducted by oxygen consumption measurements, EPR, and UV-Vis spectra. The results suggested that the protein used multiple O_2 and might be a diiron protein. All of the information could serve as a stepping stone for future studies.

All of the investigations could help us to understand how nature designs THI5p to be a single turnover protein to synthesize HMP-P without chemical and biochemical precedent.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Begley, and my committee members, Dr. Raushel, Dr. Liu, Dr. Reinhart, and Dr. Barondeau, for their guidance and support throughout the course of this research. Especially Dr. Begley, Dr. Raushel, and Dr. Liu helped me for the postdoctoral position.

During the course of this research, I received many people's help. Dr. Amrita Hazra was an excellent mentor to teach me a lot of skills. Dr. Pei-Jing Pai, Dr. Shu-Hua Chen, and Dr. Liuxi Chen in Dr. Russell's lab taught me about mass spectrometry. This knowledge was very useful when I conducted LC-MS experiments myself. Dr. Larry Dangott taught me how to conduct radioactive experiments. Darrell Martin and Yu-Shan Cheng taught me protein crystallography. Dr. Dmytro Fedoseyenko synthesized substrates.

Finally, I need to thank my friends and family, especially my wife, Dr. Anyanee Kamkaew. Without her supports, I cannot complete my PhD studies. She also gave me a precious gift, which is our son, Ryan K. Lai. We will enjoy our time together. Lastly, I thank my parents for supporting me to pursue my interest.

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

Page

Page

CHAPTER I

INTRODUCTION

Thiamin

Thiamin pyrophosphate (TPP) is an essential cofactor in all living systems [1-3]. Bacteria, fungi, and plants all possess the biosynthetic machinery to produce it, but animals must obtain it through their diets. Its biosynthesis involves two separate pathways to synthesize pyrimidine and thiazole, followed by a coupling of these two heterocycles to generate thiamin (Figure 1-1). But thiamin biosynthesis is not uniform across all species, prokaryotes and eukaryotes involve different enzymes and use different starting materials to generate the same product.

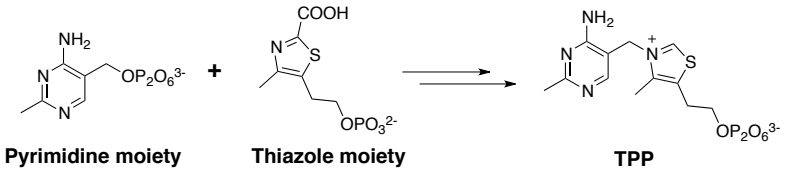


Figure 1-1. The formation of thiamin from pyrimidine and thiazole.

Biosynthesis of thiamin thiazole in prokaryotes

In bacteria, the thiazole moiety is biosynthesized by ThiG [4-6]. The early steps in thiazole biosynthesis in *Bacillus subtilis* have been studied extensively and the mechanism is proposed (Figure 1-2). The thiazole of thiamin is biosynthesized via a lysine residue of the thiazole synthase, ThiG, forming an imine with 1-deoxy-*D*-

xylulose-5-phosphate **1-1** (DXP). The imine tautomerizes to form aminoketone **1-3**. The thiocarboxylate of ThiS **1-4** attacks the ketone of **1-3** followed by an S/O acyl shift to generate **1-6** after loss of water. Subsequent elimination of ThiS forms **1-7**. The glycine imine **1-9** produced by the ThiO-catalyzed oxidation of glycine **1-8** is attacked by the thiol of **1-7** to generate **1-10**. Transimination followed by elimination of ThiG cyclizes **1-10** to afford the thiazole tautomer **1-11** [7]. The thiazole tautomer undergoes aromatization to form **1-12**.

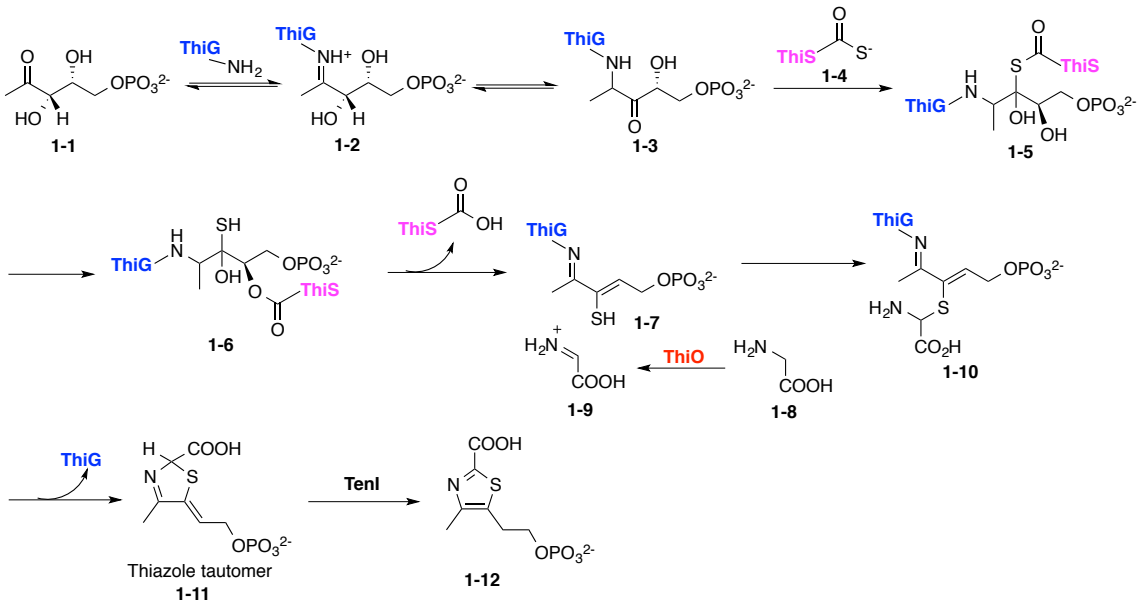


Figure 1-2. Mechanistic proposal for the formation of the thiamin thiazole in bacteria.

In chapter 2, I will discuss the reaction of the thiazole tautomer **1-11** with ThiG to undergo the reverse reaction to form **1-7**. The intermediate **1-7** was trapped by $NaBH_4$

reduction followed by SDS-PAGE analysis and identified by phosphoprotein stain and protein mass spectrometry. Furthermore, the formation of **1-7** can be used to investigate the later steps of thiazole biosynthesis involving the ThiO-catalyzed reaction. Isotope-labeled substrates were used to demonstrate the reconstitution.

Biosynthesis of thiamin pyrimidine in eukaryotes

In prokaryotes, the pyrimidine moiety, 4-amino-2-methyl-5-hydroxymethyl pyrimidine (HMP), is synthesized from 5-aminoimidazole ribonucleotide **1-13** (AIR), which is an intermediate in purine biosynthesis (Figure 1-3). The synthesis involves a complicated rearrangement that occurs via a radical reaction catalyzed by the radical SAM enzyme, ThiC [8].

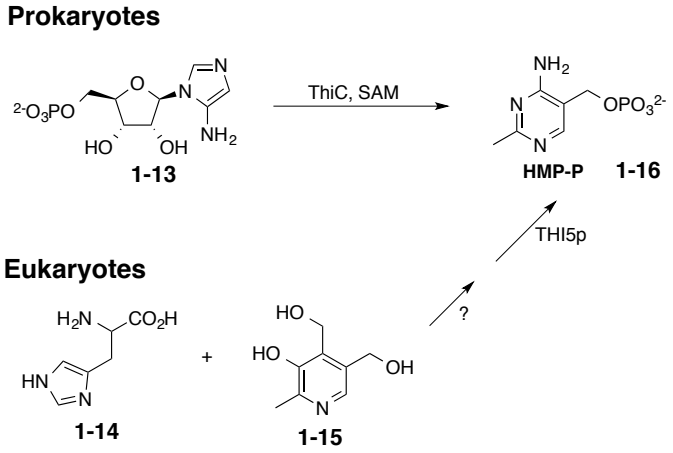


Figure 1-3. Biosynthesis of the HMP moiety of thiamin in prokaryotes and eukaryotes.

In eukaryotes, the biosynthetic pathway of the pyrimidine is still under investigation (Figure 1-3). Genetic knockout experiments in *Saccharomyces cerevisiae* have shown that the *THI5* gene family is involved in thiamin pyrimidine synthesis [9]. Feeding experiments and *in vivo* labeling studies have suggested that the pyrimidine moiety is generated from histidine **1-14** [10] and pyridoxine **1-15** [11]. The *in vivo* labeling experiments indicated C-2, C-2', N-1, C-5, C-5', and C-6 in the pyrimidine are from pyridoxine. Furthermore, the *THI5* gene is adjacent to *SNO* and *SNZ*, putative pyridoxal-5'-phosphate (PLP) biosynthetic genes [12]. Therefore, PLP could be a direct precursor for pyrimidine biosynthesis in eukaryotes. Alternatively, the addition of ^{14}N-labeled histidine in ^{15}N-growth media decreased ^{15}N-incorporation in the pyrimidine of thiamin in *S. cerevisiae* [13]. Thus, histidine is proposed to be a direct precursor for the nitrogen atom of the thiamin pyrimidine moiety (Figure 1-4). Based on feeding studies, N-1, C-2, and N-3 of the imidazole ring in histidine incorporate at N-3, C-4, and amino-N at C-4 of the pyrimidine. Specifically, the N-3 of the pyrimidine is derived from N-1 of the imidazole ring in histidine, as demonstrated by incorporation of the ^{15}N-amino group from aspartate, the origin of N-1 in histidine.

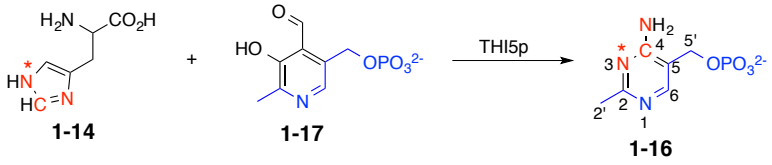


Figure 1-4. Atom incorporation patterns of histidine and PLP fragments into HMP-P.

In chapter 3 to 7, the investigation of HMP-P synthesis catalyzed by THI5p was conducted. The reconstitution of THI5p was achieved to investigate the mechanism. THI5p is a single turnover protein in which His66 in the active site reacts with PLP via the imine formation of Lys62. After HMP-P formation, His66 becomes a keto-acid with two oxygen incorporations derived from O_2. The PLP by-product is glyoxylate with oxygen incorporations. The intermediate investigations were developed and the proposed structures were made based on the LC-MS analysis. Moreover, bioinorganic investigations of THI5p were conducted to initiate future studies.

CHAPTER II

THE BACTERIAL THIAZOLE SYNTHASE:

TRAPPING OF A COVALENT INTERMEDIATE

Introduction

The biosynthesis of thiamin pyrophosphate (TPP) biosynthesis involves two separate pathways to synthesize pyrimidine and thiazole. These two heterocycles are coupled together to generate thiamin **2-14**. In *Bacillus subtilis*, the formation of thiazole is catalyzed by the thiazole synthase, ThiG, using 1-deoxy-*D*-xylulose-5-phosphate (DXP) **2-1** as starting material to react with thiocarboxylate **2-4** and glycine imine **2-9** (Figure 2-1). In this biosynthesis, the five additional enzymes, ThiF, ThiS, ThiO, ThiI, and NifS, are required to participate in this complex condensation.

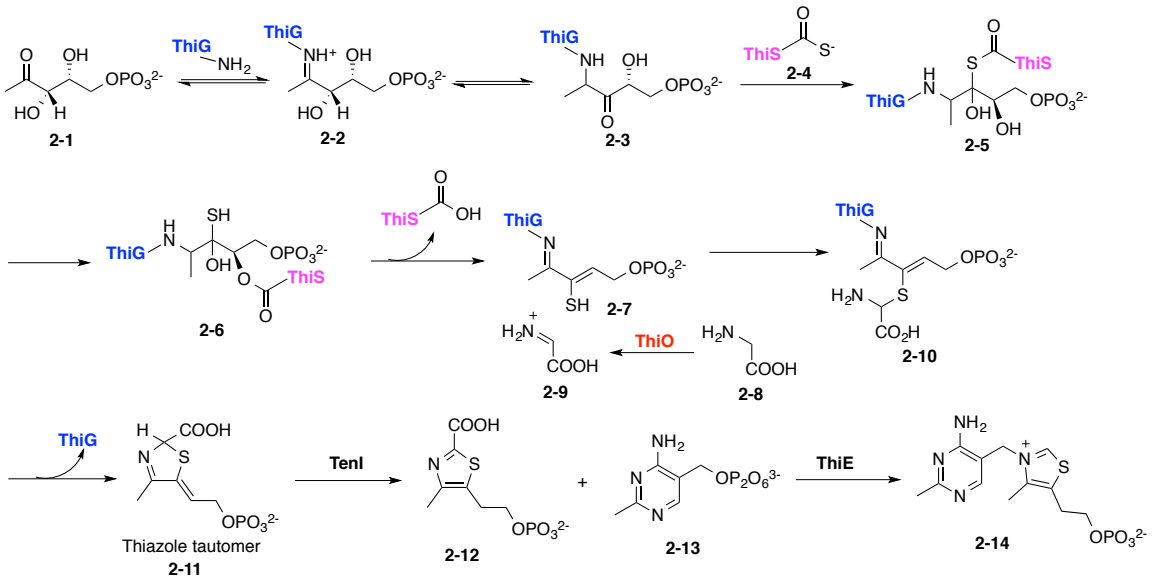


Figure 2-1. The proposed mechanism for the formation of thiamin in *B. subtilis*.

The *in vitro* reconstitution of *B. subtilis* thiazole biosynthesis was achieved. ThiG generates the thiazole tautomer **2-11** as the product. Recently, a thiazole tautomerase, TenI, has been identified to aromatize **2-11** to carboxylthiazole **2-12** [7]. Finally ThiE catalyzes the coupling reaction of **2-12** and HMP-PP **2-13** to generate thiamin. In this study, the intermediate **2-7** in thiazole biosynthesis in *B. subtilis* was identified. It was generated by the unanticipated reverse reaction of ThiG and the thiazole tautomer **2-11**. Furthermore, this intermediate **2-7** was applied to react with the glycine imine **2-9** generated by the ThiO-catalyzed glycine oxidation. The isotope-labeled glycine imine **2-9** was further used to investigate the later steps of thiazole biosynthesis involving ThiG and ThiO.

Experimental methods

Preparation of thiazole tautomer [6,14]

E. coli BL21(DE3) cells containing the overexpression plasmid (THI4p in a pET28 vector) was grown in LB medium containing kanamycin with shaking at 37 °C until OD_{600} reached 0.2 to 0.3. At this moment, THI4p overexpression was induced with IPTG (final concentration of 2 mM), and cell growth was continued at 15 °C for 16 h. The cells were harvested by centrifugation at 5000 rpm for 10 min and the resulting pellets were stored at -80 $^{\circ}$C until protein purification.

To purify THI4p, the cell pellets were resuspended in lysis buffer (10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4, pH 8) and lysed by sonication (Heat

systems Ultrasonics model, W-385 sonicator, 1.5 s cycle, 50% duty). The resulting cell lysate was clarified by centrifugation at 15000 rpm for 25 min. THI4p was purified on Ni-NTA resin following the manufacturer's (Qiagen) instructions. After elution, the protein solution was concentrated by 10 kDa MW cut-off filters (Amicon Ultra-4, Millipore).

The THI4p protein was divided into 500 µL aliquots followed by heat denaturation at 100 °C for 1.5 min. The mixtures were centrifuged at 4 °C for 10 min and the supernatant was collected followed by filtration using a 10 kDa MW cut-off filter (Amicon Ultra-4). The filtrate containing thiazole tautomer-ADP was stored in liquid nitrogen until HPLC purification.

The thiazole tautomer-ADP was purified by HPLC using the following gradient at a flow rate 2 mL/min: solvent A is water; solvent B is 100 mM NH_4OAc, pH 6.6; solvent C is methanol. 0 min: 100% B; 5 min: 10% A, 90% B; 10 min: 25% A, 60% B, 15% C; 13 min: 25% A, 60% B, 15% C; 17 min: 30% A, 10% B, 60% C; 18 min: 30% A, 10% B, 60% C; 21 min: 100% B; 30 min: 100% B. The column was Supelco LC-18-DB (250 X 4.6 mm, 5 µm ID). The thiazole tautomer-ADP was eluted at 16 min and collected. The pooled solution was lyophilized and stored in liquid nitrogen to avoid the aromatization of the thiazole tautomer-ADP.

150 µL of thiazole tautomer-ADP solution was added into 350 µL buffer (50 mM Tris-HCl and 50 mM NaCl, pH 7.6) containing 1 U of lyophilized snake venom nucleotide pyrophosphatase (EC 3.6.1.9, Sigma) and $MgCl_2$ (final concentration of 1 mM). The mixture was incubated at 37 °C for 1h followed by filtration (10 kDa MW

cut-off filter) to remove the protein. The resulting filtrate contained the thiazole tautomer and AMP. The concentration of the thiazole tautomer in the solution was calculated by determining AMP concentration by HPLC analysis.

Overexpression and purification of ThiSG

E. coli BL21(DE3) cells containing overexpression plasmid (ThiSG in pET16b vector) was grown in LB medium containing ampicillin with shaking at 37 °C. When OD_{600} reached about 0.6, ThiSG overexpression was induced by the addition of IPTG (final concentration of 0.5 mM). The cell growth was continued at 15 °C for 12 h. The cells were harvested by centrifugation and the resulting cell pellets were stored at -80 °C.

To purify ThiSG, the cell pellets were resuspended in lysis buffer (10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4, pH 8) and lysed by sonication (Heat systems Ultrasonics model, W-385 sonicator, 1.5 s cycle, 50% duty). The resulting cell lysate was clarified by centrifugation at 15000 rpm for 25 min. ThiSG was purified on a Ni-NTA resin following the manufacturer's (Qiagen) instructions. After elution, the protein solution was concentrated using 10 kDa MW cut-off filters (Amicon Ultra-4, Millipore). Finally, the protein solution was desalted by an Econo-Pac 10DG column (Bio-rad) pre-equilibrated with desalting buffer (100 mM KH_2PO_4, 100 mM NaCl, 30% glycerol, pH 7.5). The protein solution was aliquoted and stored at -80 °C.

Trapping covalent adduct by $NaBH_4$

10 µL of ThiSG (164 µM of ThiG) were incubated with two molar equivalents of thiazole tautomer in a final volume of 75 µL (50 mM Tris-HCl, 50 mM NaCl, pH 7.5) at 37 °C for 20 min. The reaction was quenched by the addition of $NaBH_4$ (0.5 mg) and incubated at RT for 30 min.

The mixture was buffer exchanged with by Amicon Ultra-0.5 three times. The buffer used for phosphoprotein stain analysis was 50 mM Tris-HCl, 50 mM NaCl, pH 7.5. The buffer for intact protein mass analysis (in Dr. Russell's Lab) was 20 mM NH_4OAc, pH 6.6.

Phosphoprotein stain analysis

The samples were prepared as previously described, however, to make the final protein concentration of every sample needed to be the same, they were loaded and separated on an SDS-PAGE gel. The gel was stained by Pro-Q Diamond Phosphoprotein Gel Stain following the manufacturer's (Invitrogen) instructions. Finally, the gel was imaged using a Typhoon scanner. After imaging, the gel was also stained with Coomassie Blue to compare the quantity of protein in each sample.

HPLC analysis of reaction reversibility

The reaction mixtures were directly filtered using a 10 kDa MW cut-off filter (VWR) without the addition of $NaBH_4$. The filtrates were analyzed by HPLC using the following linear gradient at a flow rate 1 mL/min over a Supelco LC-18-DB (250 X 4.6

mm, 5 μm ID) column: solvent A is water, solvent B is 100 mM K_2HPO_4, pH 6.6, solvent C is methanol. 0 min: 100% B; 5 min: 10% A, 90% B; 10 min: 25% A, 60% B, 15% C; 13 min: 25% A, 60% B, 15% C; 17 min: 30% A, 10% B, 60% C; 18 min: 30% A, 10% B, 60% C; 21 min: 100%B.

Isotopic incorporation without ThiO

170 μL of 200 μM of thiazole tautomer was incubated with 100 μL of 220 μM of ThiSG in a final volume of 500 μL (50 mM Tris-HCl, 50 mM NaCl, pH 7.5) at 37 °C for 20 min. There were a total of six reaction mixtures. After the reaction, the reaction mixtures were collected followed by desalting with an Econo-Pac 10DG column pre-equilibrated with 50 mM Tris-HCl, 50 mM NaCl, pH 7.5. The desalted protein solution was collected and separated into seven 500 μL aliquots. Then, 80 μL of 100 mM $^{15}NH_4Cl$ and glyoxylate (50 mM Tris-HCl, 50 mM NaCl, pH 7.5) were added to each. The reaction mixtures were incubated at 37 °C for 1 h, followed by removal of the proteins with a 10 kDa MW cut-off filter.

The filtrates were analyzed by HPLC. The labeled thiazole tautomer was collected from 7.2 to 8 min by HPLC following the same procedure as the purification of the thiazole tautomer-ADP. After collection, the sample was lyophilized for 1 week. The degraded product, CO_2-Vinyl-Thz, was isolated at 25 min and analyzed with negative-ESI.

Isotopic incorporation with ThiO

100 μL of 220 μM ThiO was incubated with 80 μL of 100 mM $^{15}NH_4Cl$ and 1,2-^{13}C-glycine at room temperature for 1 h. At the same time, 170 μL of 200 μM thiazole tautomer was incubated with 100 μL of 220 μM ThiSG in final volume of 500 μL (50 mM Tris-HCl, 50 mM NaCl, pH 7.5) at 37 °C for 20 min. There were a total of six reaction mixtures. After the reaction, the reaction mixtures were collected followed by desalting over an Econo-Pac 10DG column pre-equilibrated with 50 mM Tris-HCl, 50 mM NaCl, pH 7.5. The desalted protein solution was collected and divided into seven 500 μL aliquots.

Then ThiO reaction mixtures were added to each desalted ThiSG solution and incubated at 37 °C for 1 h, and then filtered using a 10 kDa MW cut-off filter. The filtrates were analyzed by HPLC. The labeled thiazole tautomer was collected from 7.2 to 8 min by HPLC using the same method as the purification of the thiazole tautomer-ADP. After collection, the sample was lyophilized for 1 week. The degraded product, CO_2-Vinyl-Thz, was isolated at 25 min and analyzed with negative ESI.

Results and discussion

During investigating TenI, the thiazole tautomer **2-11** was incubated with TenI and ThiG (The experiments were conducted by Dr. Amrita Hazra). The HPLC analysis showed the disappearance of **2-11**, but no formation of carboxythiazole **2-12** (Figure 2-2A and 2-2B).

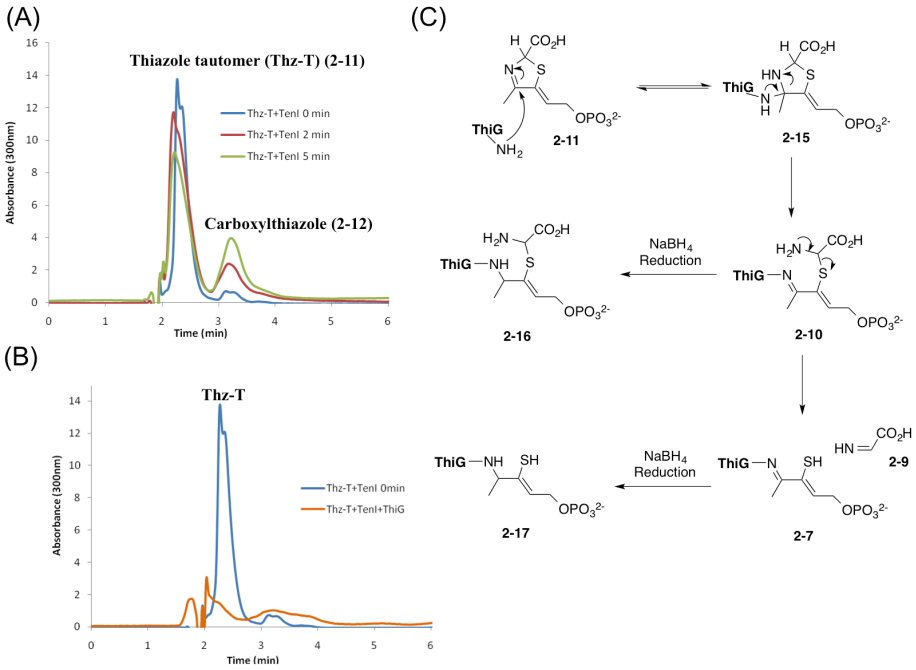


Figure 2-2. (A) HPLC analysis of TenI assays showed TenI can convert the thiazole tautomer **2-11** to carboxythiazole **2-12**. (B) HPLC analysis of the thiazole tautomer **2-11** incubated with TenI and ThiG assay showed that there was no formation of carboxythiazole **2-12** and the thiazole tautomer **2-11** was consumed. (C) The proposed mechanism of the ThiG reverse reaction for the formation of **2-7** and **2-10**. $NaBH_4$ reduces the imine of **2-7** and **2-10** to generate a possibly stable **2-16** and **2-17** for identification. The experiments were conducted by Dr. Amrita Hazra.

This result led us to investigate the possibility of a ThiG reverse reaction with the thiazole tautomer **2-11**. During thiazole biosynthesis, ThiG uses the active site lysine to form an imine with **2-1** [15]. It was proposed that during the ThiG reverse reaction the

active-site lysine reacts with the thiazole tautomer **2-11** followed by ring opening to generate **2-10** (Figure 2-2C). It could further undergo elimination of the glycine imine to form **2-7**. To test our hypothesis, the thiazole tautomer **2-11** was incubated with ThiG to investigate the possibility of the reverse reaction. The thiazole tautomer **2-11** was reported to be generated by thiazole tautomer-ADP, which co-purified with *Saccharomyces cerevisiae* thiazole synthase, after nucleotide pyrophosphatase treatment.

Since intermediates **2-7** and **2-10** (Figure 2-2C) have a phosphate and an imine, the full reaction assay (ThiSG + thiazole tautomer) was treated with $NaBH_4$ to reduce the imine and form a stable intermediate. The samples were analyzed by SDS-PAGE with ProQ Diamond phosphoprotein stain (Invitrogen) to detect phosphorylated protein (Figure 2-3). After comparing with control assays, the full reaction assay gave the most intense ThiG signal. The full reaction without $NaBH_4$ treatment showed a faint background signal. The background signal cannot be explained because the company refused to reveal the ingredients. But the results suggested the presence of phosphate and imine on ThiG after the reaction. Furthermore, there was no ThiG signal difference in all of assays when the active-site lysine was mutated. Lastly, thiazole tautomer-ADP containing the thiazole tautomer moiety was tested with ThiSG. The phosphoprotein stain analysis suggested that ThiG did not react with it. These results implied that ThiG reacts with the thiazole tautomer **2-11** to form an intermediate with the phosphate and imine group.

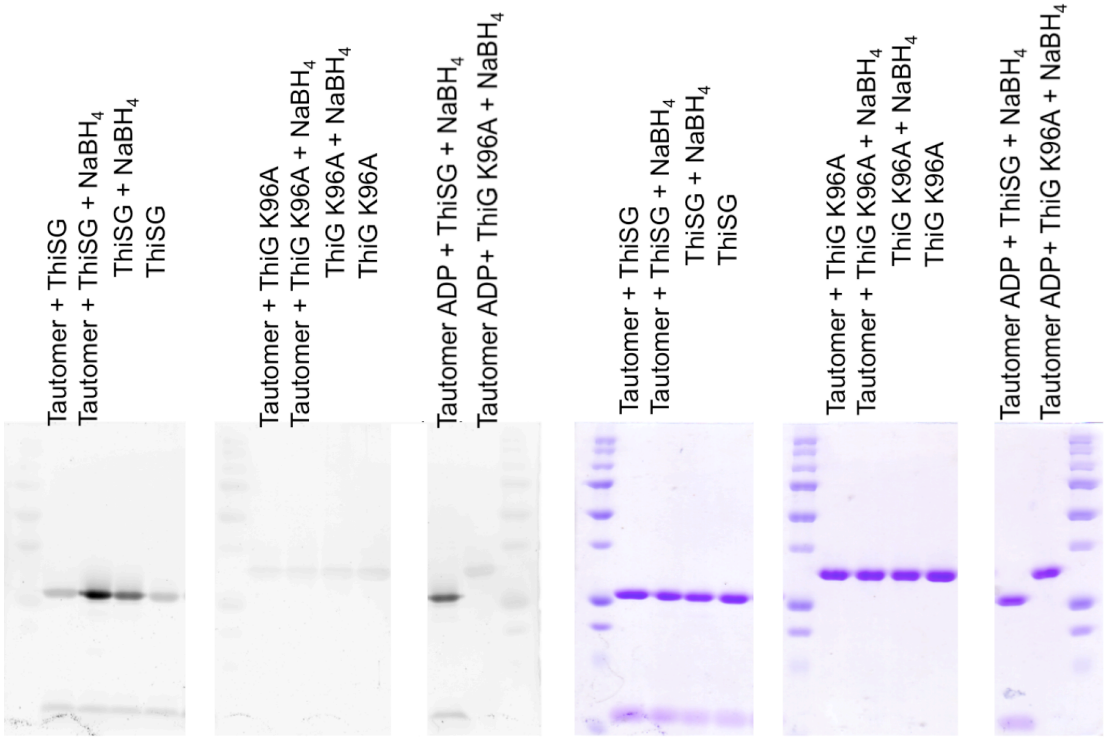


Figure 2-3. ThiSG and ThiG K96A reactions analyzed by SDS-PAGE with phosphoprotein stain (Left) and Coomassie Blue (Right). The MW difference of ThiG in ThiSG (in pET16b) and ThiG K96A (in pET28b) were due to different His-tag sequences.

In order to characterize the intermediate, the full reaction sample after $NaBH_4$ treatment was analyzed by mass spectrometry. After comparing with the control sample, the molecular weight (MW) of ThiG in the full reaction assay increased by 196 Da. Based on the proposed reverse reaction (Figure 2-2C), +196 Da corresponds to **2-17**, which is derived from **2-7** after $NaBH_4$ reduction (Figure 2-4A). Furthermore, the full

reaction assay was carried out with ThiG K96A following the same procedure. ThiG K96A did not show a MW shift (Figure 2-4B). The results tallied with phosphoprotein stain analysis and suggest that the active site lysine reacts with the thiazole tautomer to undergo the reverse reaction.

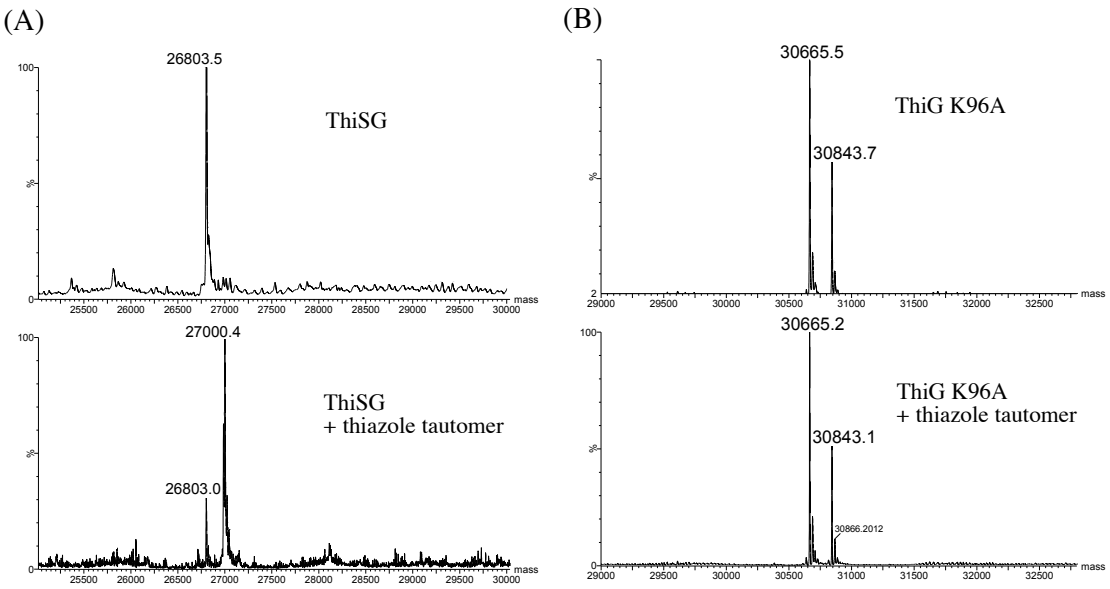


Figure 2-4. The MW measurement of ThiSG and ThiG K96A after $NaBH_4$ treatment. (A) ThiG increased by 196 Da in the full reaction assay. (B) ThiG K96A was not modified in the full reaction assay. The MW difference of ThiG in ThiSG (in pET16b) and ThiG K96A (in pET28b) were because of different His-tag sequences.

In the proposed thiazole biosynthesis (Figure 2-1), intermediate **2-7** reacts with the glycine imine **2-9** generated by ThiO-catalyzed glycine oxidation to generate the thiazole tautomer **2-11**. In our previous studies, we have shown successful reconstitution

using ThiG, ThiS, NifS, ThiF, and ThiO. But the identification of intermediate **2-7** could allow us to investigate the later steps of thiazole biosynthesis involving ThiO and ThiG only. Therefore, the competence of intermediate **2-7** in this biosynthesis was investigated to observe the regeneration of the thiazole tautomer **2-11** after the addition of glycine imine **2-9**. Therefore, intermediate **2-7** was generated by the reverse reaction followed by desalting to remove excess thiazole tautomer **2-11**. The resulting solution containing intermediate **2-7** was incubated with glycine imine **2-9** generated by glyoxylate and ammonium chloride. Finally, $NaBH_4$ was added to quench the reaction and observe the regeneration of unmodified ThiG by mass spectrometry (Figure 2-5A). In addition, the assays without $NaBH_4$ treatment were analyzed by HPLC to monitor the quantity of the thiazole tautomer **2-11** (Figure 2-5B). In the reverse reaction assay, ThiSG was incubated with two molar equivalents of the thiazole tautomer **2-11**. The quantity of **2-11** decreased after the reaction. The addition of glycine imine **2-9** into the reverse reaction assay showed the regeneration of the thiazole tautomer **2-11**. These results suggested that the intermediate **2-7** could react with the glycine imine **2-9** to regenerate the thiazole tautomer **2-11** and demonstrated the competency of the intermediate **2-7**.

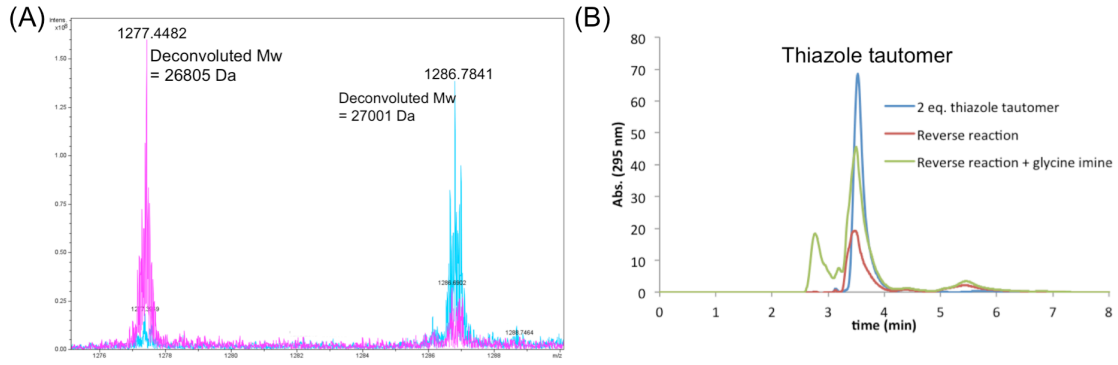


Figure 2-5. (A) The overlapping $[M+21H]^{21+}$ spectra of the reverse (blue) and forward (pink) reactions suggested that the addition of glycine imine **2-9** in the forward reaction to the intermediate **2-7** from the reverse reaction regenerated unmodified ThiG. (B) The HPLC analysis of the ThiG assays showed that the regeneration of the thiazole tautomer **2-11** in the reverse reaction with the addition of the glycine imine **2-9** compared with the reverse reaction.

To investigate the identity of the intermediate **2-7**, different isotope-labeled glycine imines **2-9** were used to generate the isotope-labeled thiazole tautomer **2-11**. However, the thiazole tautomer **2-11** was prone to degradation, so the thiazole tautomer **2-11** was lyophilized to degrade as carboxylvinylthiazole **2-18** for MS analysis (Figure 2-6). First, $^{15}NH_4Cl$ was incubated with glyoxylate to generate ^{15}N-glycine imine. ^{15}N-glycine imine was added into the solution of intermediate **7**. The MW of carboxylvinylthiazole **2-18** had +1 Da (Figure 2-6B). Furthermore, 1,2-^{13}C-glycine imine could be generated via the oxidation of 1,2-^{13}C-glycine catalyzed by ThiO. In this experiment, carboxylvinylthiazole **2-18** had +2 Da (Figure 2-6C). These experiments

demonstrated the competency of the intermediate **2-7** and the reconstitution of the later steps of thiazole biosynthesis involving ThiG and ThiO. Lastly, the possible substrate channel between ThiG and ThiO was studied by adding an excess of $^{15}NH_4Cl$ to the ThiO-catalyzed 1,2-^{13}C-glycine oxidation reaction. The result showed +2 and +3 Da indicating that $^{15}NH_4Cl$ reacted with ^{13}C-glycine imine to generate ^{13}C,^{15}N-glycine imine (Figure 2-6D). This suggested that there is no compact channel between ThiO and ThiG.

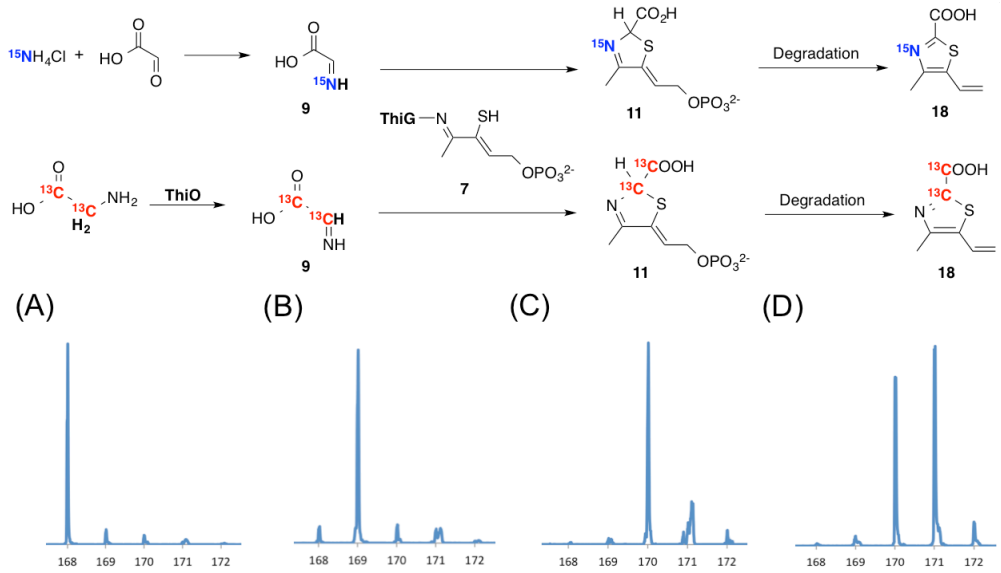


Figure 2-6. Mass spectra of carboxylvinylthiazole **2-18** suggesting isotope incorporation. (A) Carboxylvinylthiazole standard; (B) ^{15}N-glycine imine generated by $^{15}NH_4Cl$ and glyoxylate; (C) 1,2-^{13}C-glycine imine generated by oxidation of 1,2-^{13}C-glycine catalyzed by ThiO; (D) $^{15}NH_4Cl$ and 1,2-^{13}C-glycine with ThiO.

Conclusion

This chapter has demonstrated that the intermediate **2-7** in thiazole biosynthesis was generated by the reverse reaction. Moreover, the intermediate **2-7** was used to investigate thiazole biosynthesis involving ThiG and ThiO in the later steps.

CHAPTER III

THIAMIN PYRIMIDINE BIOSYNTHESIS IN *CANDIDA ALBICANS*: A REMARKABLE REACTION BETWEEN HISTIDINE AND PYRIDOXAL PHOSPHATE[*]

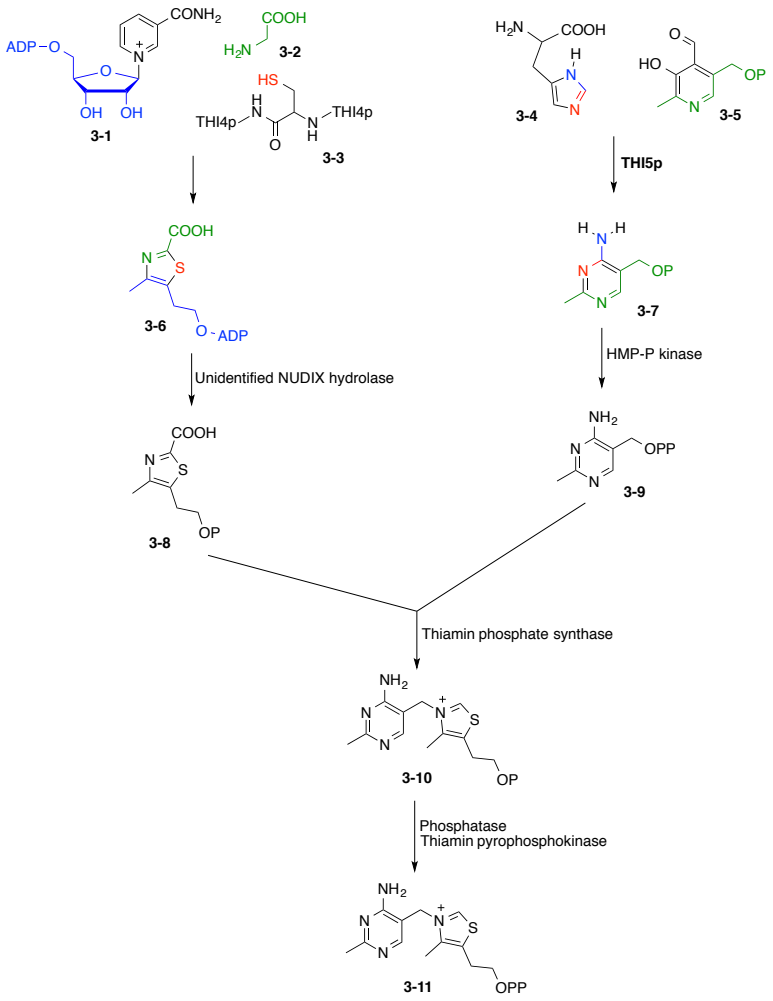
Introduction

In *Saccharomyces cerevisiae*, thiamin is biosynthesized from thiazole (**3-8**) and pyrimidine (**3-9**) moieties. Thiazole **3-8** has been investigated to show that the thiazole synthase (THI4p) uses an active-site cysteine, glycine and nicotinamide adenine dinucleotide (NAD) to make thiazole moiety [16].

The past feeding experiments implied 4-amino-2-methyl-5-hydoxymethylpyrimidine phosphate (HMP-P, **3-7**) is formed from histidine (**3-4**) and pyridoxal 5'-phosphate (PLP, **3-5**) [11,17-19]. Furthermore, genetic studies have demonstrated that HMP-P formation requires only a single gene (*THI5*) [9]. Based on all information, HMP-P is formed using remarkable chemistry without any chemical or biochemical precedent. In this chapter, the overexpression of THI5p, the reconstitution of the enzyme activity, and preliminary characterization of the reaction are reported. These experiments suggest that the THI5 protein is the histidine source for HMP-P formation and that THI5p is a single turnover protein.

---

[*]Reprinted with permission from "Thiamin Pyrimidine Biosynthesis in Candida albicans: A Remarkable Reaction between Histidine and Pyridoxal Phosphate" by Lai, R.-Y.; Huang, S.; Fenwick, M. K.; Hazra, A.; Zhang, Y.; Rajashankar, K.; Philmus, B.; Kinsland, C.; Sanders, J. M.; Ealick, S. E.; Begley, T. P., 2012. *Journal of the American Chemical Society*, 134, 9157-9159, Copyright [2012] by American Chemical Society.



Scheme 3-1. Thiamin pyrophosphate biosynthesis in *S. cerevisiae* [20].

Experimental methods

Source of chemicals

All chemicals were purchased from Sigma-Aldrich Corporation (U.S.A.) unless otherwise mentioned. Trypsin was obtained from Promega. M9 salts were obtained from Becton Dickinson. Kanamycin and IPTG were purchased from LabScientific Inc.

Analytical HPLC (Agilent 1260 instrument) was carried out using a Supelcosil LC-18 column (250 mm X 10 mm, 5 μm ID). LC-MS (Agilent 1260 instrument and Bruker microTOF-Q II) was carried out using a Supelcosil LC-18-T column (150 mm X 3 mm, 3 μm ID). The solvents for HPLC and LC-MS were HPLC grade and LC-MS grade respectively and were obtained from EMD. Sequencing grade trypsin was obtained from Promega. Slide-A-Lyzer Dialysis Cassettes, 10K MWCO were purchased from Thermo Scientific.

THI5p overexpression and purification

E. coli BL21(DE3) cells containing the *THI5* gene in pET28b was grown in minimal medium (11.3 g M9 salts, 13.3 mL 50% glucose, 2.7 mL 1 M $MgSO_4$, 100 μL 1 M $CaCl_2$ diluted to 1 L medium) containing kanamycin (40 μg/mL) with shaking at 37 °C until the OD_{600} reached 0.6. At this point, protein overexpression was induced with IPTG (final concentration, 500 μM), and cell growth was continued at 15 °C for 16 h. The cells were harvested by centrifugation, and the cell pellets from 1 L of culture were resuspended in 20 mL of lysis buffer (10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4, 5 mM DTT, pH 8.0) and lysed by sonication (Heat System Ultrasonics model W-385 sonicator, 1.5 s cycle, 60% duty). The resulting cell lysate was clarified by centrifugation and the THI5 protein was purified on a Ni-NTA column following the manufacturer's instructions. After elution, the protein was desalted under anaerobic conditions using a 10DG column (BioRad) pre-equilibrated with 100 mM Tris-HCl

buffer, 5 mM DTT, 30% glycerol, pH 7.5. The purified protein was stored in aliquots in liquid nitrogen. The purity was analyzed by SDS PAGE and ESI-MS (Figure 3-1).

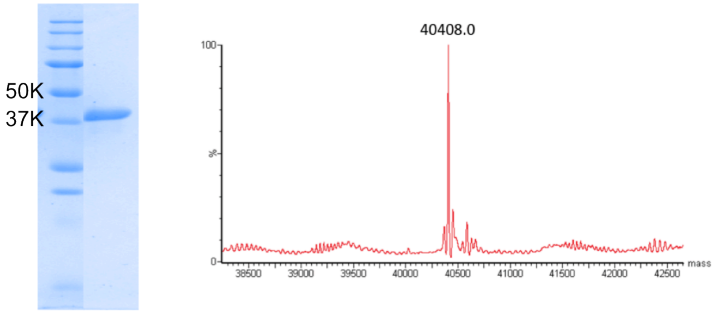


Figure 3-1. Analysis of purified THI5p. A) SDS PAGE analysis. B) ESI-MS analysis [20].

Reconstitution of the THI5p activity

100 µL of 600 µM THI5p was anaerobically pre-incubated with 1.5 equivalent of $Fe(NH_4)_2(SO_4)_2$ in an ice bath for 30 min. The mixture was anaerobically desalted using a Bio-Spin 6 column (Bio-Rad) pre-equilibrated with 100 mM Tris-HCl, 5 mM DTT, pH 7.5. Subsequently, 2 molar equivalents of PLP were added. The reaction mixture was aerobically incubated at room temperature for 3 h, filtered using a 10 kDa MW cut-off filter, and analyzed by HPLC and LC-MS.

HPLC conditions for analyzing the THI5p reaction mixture

The following linear gradient, at a flow rate of 2 mL/min, on a Supelcosil LC-18 column (250 mm X 10 mm, 5 µm ID) was used: solvent A is water, solvent B is 100

mM KPi, pH 6.6, solvent C is methanol; 0 min: 100% B; 5 min: 10% A, 90% B; 10 min: 25% A, 60% B, 15% C; 13 min: 25% A, 60% B, 15% C; 17 min: 30% A, 10% B, 60% C; 18 min: 30% A, 10% B, 60% C; 21 min: 100% B.

^{15}N-THI5p overexpression and purification

The overexpression and purification procedures were the same as described above for THI5p except $^{14}NH_4Cl$ was replaced by $^{15}NH_4Cl$ (Cambridge Isotope Laboratories, Inc.) in the M9 salts. Analysis of the labeled and unlabeled proteins by ESI-MS confirmed a high level of ^{15}N incorporation in the ^{15}N-THI5p (Figure 3-2).

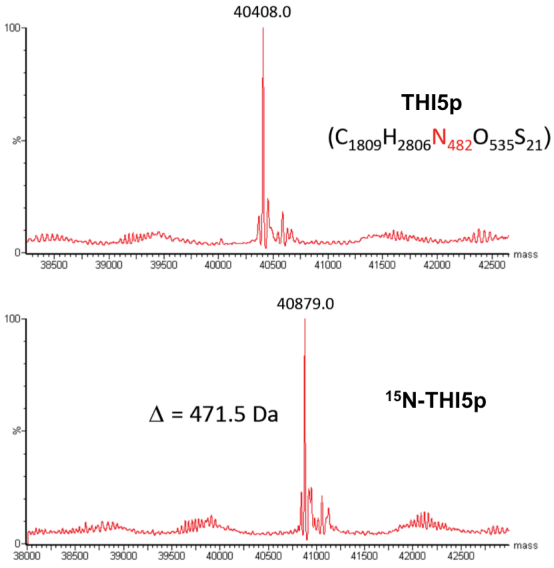


Figure 3-2. The ESI-MS analysis of ^{14}N-THI5p and ^{15}N-THI5p [20].

LC-MS analysis of the THI5p reaction mixture

The following linear gradient, at a flow rate of 0.4 mL/min, on a Supelcosil LC-18-T column (150 mm X 3 mm, 3 μm ID) was used: solvent A is 20 mM NH_4OAc, pH 6.6, solvent B is 75% methanol; 0 min: 100% A; 3 min: 100% A; 8 min: 50% A, 50% B; 10 min: 100% A. The HPLC is coupled with a microTOF-Q II for MS analysis in the positive mode.

Trypsin digestion of inactive THI5p

After completion of HMP-P formation, the inactive THI5p (50 μg, 1.8 μL of 660 μM THI5p) was added to 10 μL of 6 M guanidine-HCl, 25 mM ammonium bicarbonate, pH 8.0. 1 μL of 200 mM DTT in 25 mM ammonium bicarbonate, pH 8.0, was then added to the denatured protein solution and the resulting mixture was incubated at room temperature for 1 hour. After that, 10 μL of 200 mM iodoacetamide in 25 mM ammonium bicarbonate pH 8.0, were added and the reaction mixture was incubated at room temperature in the dark for 1 hour. Finally, 77.5 μL of 25 mM ammonium bicarbonate, pH 8.0, were added to reduce the guanidine-HCl concentration to 0.6 M followed by the addition of 1 μg of trypsin. The mixture was incubated at 37 °C overnight and stored at -20 °C until analysis.

LC-MS conditions for analyzing THI5p samples after trypsin digestion

The following linear gradient, at a flow rate of 0.4 mL/min, on a Supelcosil LC-18-T column (150 mm X 3 mm, 3 μm ID) was used: solvent A is 0.1% formic acid in

water, solvent B is 0.1% formic acid in acetonitrile; 0 min: 95% A, 5% B; 1 min: 100% A; 60 min: 35% A, 65% B. The HPLC is coupled with a microTOF-Q II for MS analysis in the positive mode.

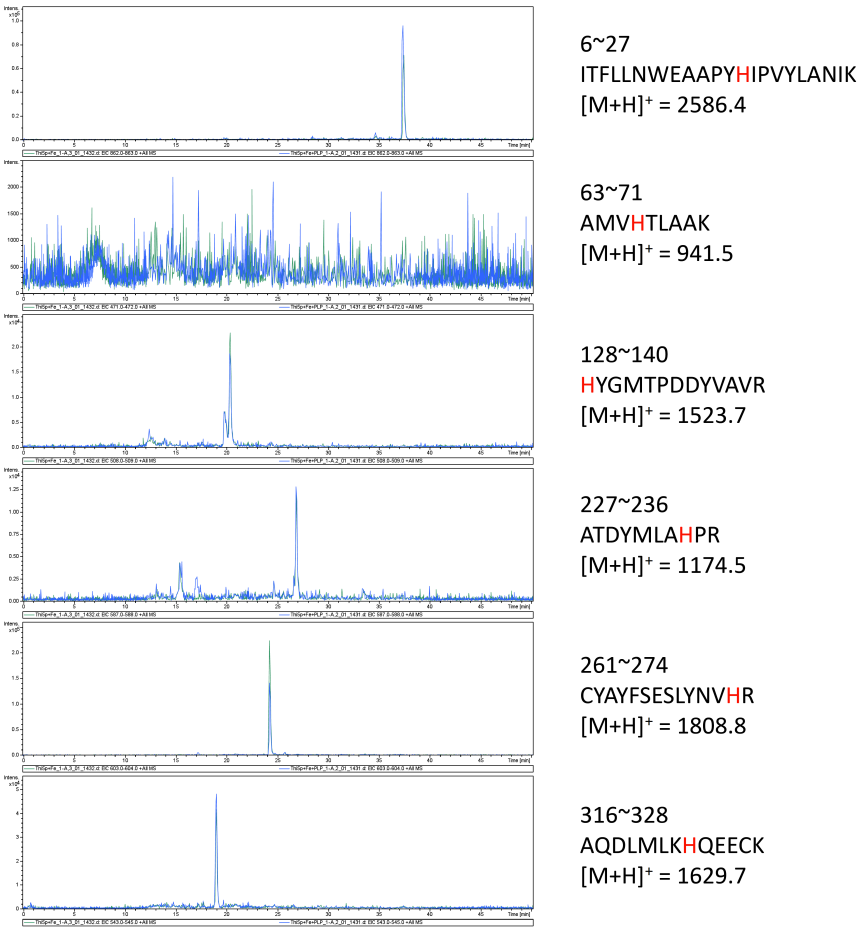


Figure 3-3. LC-MS analysis (extracted ion chromatograms) of the histidine-containing tryptic peptides derived from active and inactive THI5p [20]. Blue trace: peptides derived by trypsinolysis from the THI5p+Fe(III)+PLP reaction mixture. Green trace: peptides derived by trypsinolysis from the THI5p+Fe(III) control. All expected histidine containing peptides were detected except for His66.

THI5p refolding experiments

1 mL of 700 μM THI5p was added to 25 mL of denaturing buffer (50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 8 M urea, pH 7.8) and incubated at room temperature for 1 hour. The solution was then transferred to a Slide-A-Lyzer Dialysis Cassette, 10K MWCO (Thermo Scientific) and dialyzed against 2 L of dialysis buffer (50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 5 M urea, pH 7.8) for 5 h. Subsequently, the cassette was transferred to the next buffer (50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 2 M urea, pH 7.8) and dialyzed overnight. The protein solution was then loaded into a 5 mL Ni-NTA column pre-equilibrated with 50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 2 M urea, pH 7.8. The Ni-NTA column was installed on an FPLC and eluted, at a flow rate of 1 mL/min, with the following linear gradient: solvent A is 50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 2 M urea, pH 7.8; solvent B is 50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, pH 7.8; 0 min: 100% A; 120 min: 100% B. Finally, the refolding protein was eluted by 200 mM imidazole, 50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, pH 7.8. After concentration, the protein solution was buffer-exchanged using a Bio-Spin 6 Column (Bio-Rad) into 100 mM Tris-HCl, 5 mM DTT, pH 7.5. The activity of the refolded THI5p protein was determined as described above for the native protein.

Mutagenesis of THI5p

The mutants shown in Table 3-1 were constructed. The activity of each mutant was determined as described above for the native protein. Most of the H18A mutant overexpressed as inclusion bodies, however a very small amount of this mutant could be purified from the crude lysate and was found to be active.

Table 3-1. THI5p histidine mutants and their activities [20]

Mutant		Mutant	
H18N	Inclusion body	H18D	Inclusion body
H66N	Inactive	H18A	Inclusion body
H128N	Active	H66G	Inactive
H234N	Active		
H273N	Active		
H323N	Active		

Results and discussion

Reconstitution of protein activity

Candida albicans THI5p was heterologously overexpressed in *E. coli* BL21 (DE3) cells grown in minimal media. The protein was purified by Ni-NTA chromatography. Finally the protein was buffer-exchanged anaerobically and stored in liquid nitrogen until the further studies.

The active THI5p was incubated with Fe(II) or Fe(III) anaerobically followed by removal of excess iron. Addition of PLP to the protein solution under aerobic condition generated HMP-P. The product was characterized by NMR and ESI-MS spectrometry.

The HPLC analysis showed the reaction was metal dependent (Figure 3-4). In the presence of EDTA, the production of HMP-P was abolished. It also required aerobic condition to generate HMP-P. Moreover, the addition of histidine did not affect the amount of HMP-P production.

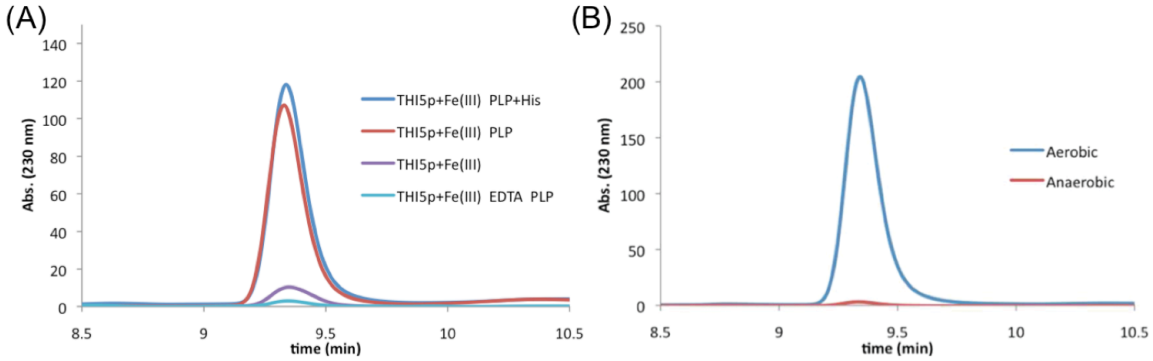


Figure 3-4. Reconstitution of the THI5-catalyzed reaction [20]. (A) The reconstitution reaction requires PLP and Fe(III) but is independent of added histidine. The small amount of HMP-P shown in the green trace is due to product that co-purifies with THI5 and not to synthesis. (B) The reaction to form HMP-P requires oxygen.

This result seemed not tally with the past feeding experiments. Hence, we hypothesized that one of histidine residues could be the N=C-N donor for HMP-P formation. In order to test this hypothesis, ^{15}N-THI5p was prepared by growing

overexpression strain in $^{15}NH_4Cl$-contiaining minimal medium. ^{15}N-THI5p was used to set up the assays analyzed by LC-MS (Figure 3-5). The molecular weight of HMP-P showed the expected [M+2] peak corresponding to incorporation of two ^{15}N-atoms from protein. This suggested that THI5p serves as the donor of the histidine-derived atoms of HMP-P.

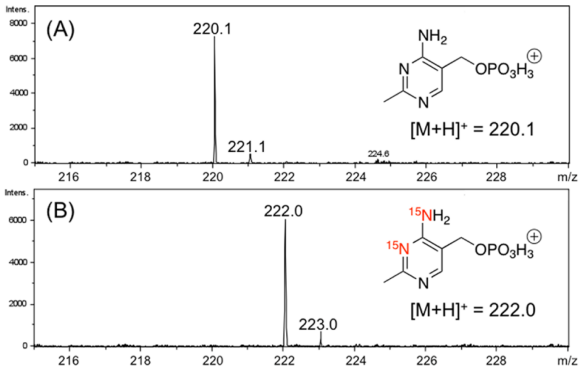


Figure 3-5. MS analysis of HMP-P formed using (A) ^{14}N-THI5p and (B) ^{15}N-THI5p [20].

THI5p refolding experiment

Although ^{15}N-THI5p assays gave an intriguing result for possibility of THI5p histidine residue as the substrate, it could be a co-purified metabolite served as the substrate. In order to eliminate this concern, THI5p was denatured in 8 M urea solution followed by slow dialysis to 2 M urea solution. Then the THI5p solution was loaded into Ni-NTA column. The denatured THI5p could be refolded by gradually decreasing urea concentration by FPLC. Finally, THI5p could be eluted and concentrated to test its

activity. The refolding THI5p still showed about 59% of the specific activity of the original preparation (Figure 3-6).

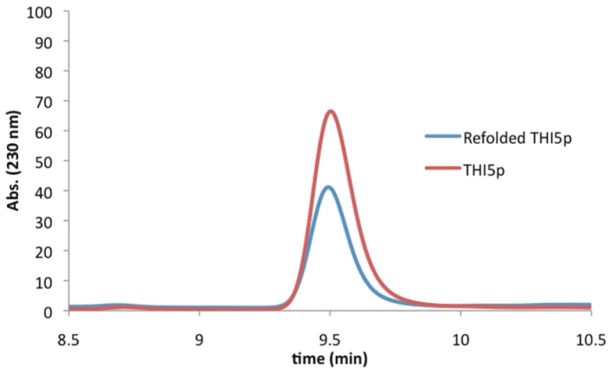


Figure 3-6. THI5p denatured in urea and refolded is active [20]. Red trace: HMP-P produced by THI5p before urea denaturation. Blue trace: HMP-P produced by THI5p after denaturation in 8 M urea and renaturation by slow dialysis.

Since denaturation destroys the active-site structure and dialysis removes any released metabolite, this experiment provides strong support for the hypothesis that the histidine-derived N=C−N fragment is extracted from the THI5 protein.

Trypsin digestion of inactive THI5p and mutagenesis of THI5p

All of preliminary results indicated that THI5p serves as N=C-N donor for HMP-P biosynthesis. But we do not know which residue as the substrate. Since the past

feeding experiments suggested histidine as the substrate, there are six histidine residues in THI5p as the candidates to test.

Hence the inactive THI5p after full reaction assay was digested by trypsin followed by LC-MS analysis. The extracted ion chromatography (EIC) was used to see the disappearance of the peptides containing each histidine (Figure 3-3). Five of six peptides could be found but without disappearance compared with apo-THI5p digestion sample. Unfortunately the peptide containing His66 cannot be identified in both of samples.

In order to get more information, all of histidine mutants were overexpressed and test their activities. His18 mutants formed the inclusion body, but fortunately H18A could be purified as a very tiny amount. H18A was active to generate HMP-P. In other five histidine mutants, only His66 was inactive to generate HMP-P after mutation.

With these two experiments, they implied that His66 might be the N=C-N donor to form HMP-P.

Active site of THI5p crystal structure

This work was done by Dr. Siyu Huang and Michael K. Fenwick at Dr. Steve E. Ealick lab at Cornell University. The active site of THI5p showed PLP forms imine with Lys62 and His66 is closed to PLP (Figure 3-7).

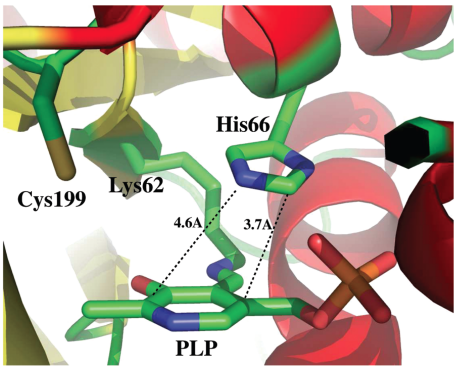


Figure 3-7. Active site of *C. albicans* THI5p showing PLP bound via an imine to Lys62 and His66 in close proximity to the PLP [20]. This was done by Dr. Siyu Huang.

The crystal structure was complementary with the results of mutagenesis of THI5p. Based on all of results, His66 is the N=C-N donor for HMP-P synthesis.

Conclusion

C. albicans THI5p can be readily overexpressed in *E. coli*. Furthermore the protein uses PLP and the active-site His66 to form HMP-P. The reaction also requires oxygen and iron.

CHAPTER IV

CHARACTERIZATION OF THE MODIFIED HIS66 AND PLP BY-PRODUCTS

Introduction

In *Saccharomyces cerevisiae* thiamin biosynthesis (Scheme 3-1), the thiazole synthase (THI4p) and the pyrimidine synthase (THI5p) have been investigated as single turnover proteins [16,20]. In thiazole biosynthesis, THI4p uses an active-site cysteine to react with glycine and nicotinamide adenine dinucleotide (NAD) to generate the thiazole moiety. In pyrimidine biosynthesis, His66 **4-1** in the THI5p active-site reacts with pyridoxal 5'-phosphate **4-2** (PLP) via an imine bound with Lys62 to form 4-amino-2-methyl-5-hydoxymethylpyrimidine phosphate **4-3** (HMP-P). The reaction requires both oxygen and iron. This biosynthesis involves remarkable chemistry without any precedent. To study the reaction mechanism, the characterization of His66 in inactive THI5p (after HMP-P formation) is necessary, as well as identification of the by-product of PLP **4-2**. In this chapter, the chemical equation of HMP-P **4-3** formation is solved for further mechanistic investigations.

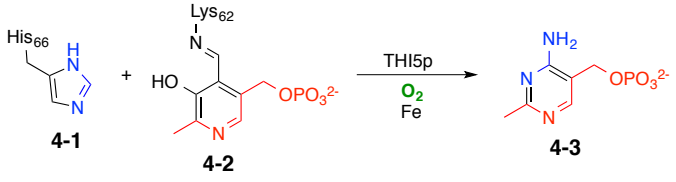


Figure 4-1. The result of the THI5p reconstitution in chapter 3.

Experimental methods

Source of chemicals

All chemicals were purchased from Sigma-Aldrich (U.S.A.) unless otherwise noted. M9 salts were obtained from Becton Dickinson. Kanamycin and IPTG were purchased from LabScientific Inc. Analytical HPLC (Agilent 1260 instrument) was carried out using a Supelcosil LC-18 column (250 mm X 10 mm, 5 μm ID). LC-MS (Agilent 1260 instrument and Bruker microTOF-Q II) was carried out using a Supelcosil LC-18-T column (150 mm X 3 mm, 3 μm ID) and Synergi Polar-RP 100A column (50 mm X 2 mm, 2.5 μm). The solvents for HPLC and LC-MS were HPLC grade and LC-MS grade respectively and were obtained from EMD. Sequencing grade trypsin was obtained from Promega.

THI5p overexpression and purification

E. coli BL21(DE3) cells containing the *THI5* gene in pET28b vector was grown in minimal media (11.3 g M9 salts, 13.3 mL 50% glucose, 2.7 mL 1 M $MgSO_4$, 100 μL 1 M $CaCl_2$ diluted to 1 L medium) containing kanamycin (40 μg/mL) with shaking at 37 °C until the OD_{600} reached 0.6. Protein overexpression was then induced with IPTG (final concentration of 500 μM), and cell growth was continued at 15 °C for 16 h. The cells were harvested by centrifugation, and the cell pellets from 1 L of culture were resuspended in 20 mL of lysis buffer (10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4, 5 mM DTT, pH 8.0) and lysed by sonication (Heat System Ultrasonics model W-385 sonicator, 1.5 s cycle, 60% duty). The resulting cell lysate was clarified by

centrifugation and the THI5 protein was purified on a Ni-NTA column following the manufacturer's instructions. After elution, the protein was desalted under anaerobic conditions using an Econo-Pac 10DG column (BioRad) pre-equilibrated with 100 mM Tris-HCl buffer, 5 mM DTT, 30% glycerol, pH 7.5. The purified protein was stored in aliquots in liquid nitrogen.

Reconstitution of the THI5p activity

100 μL of 600 μM THI5p was anaerobically pre-incubated with 1.5 molar equivalent of $Fe(NH_4)_2(SO_4)_2$ in an ice bath for 30 min. The mixture was anaerobically desalted using a Bio-Spin 6 column (Bio-Rad) pre-equilibrated with 100 mM Tris-HCl, 5 mM DTT, pH 7.5. Subsequently, 2 equivalents of PLP were added. The reaction mixture was aerobically incubated at room temperature for 3 h, filtered using a 10 kDa MW cut-off filter, and analyzed by HPLC and LC-MS.

HPLC conditions for analyzing the THI5p reaction mixture

The following linear gradient, at a flow rate of 2 mL/min, on a Supelcosil LC-18 column (250 mm X 10 mm, 5 μm ID) was used: solvent A is water, solvent B is 100 mM KH_2PO_4, pH 6.6, solvent C is methanol; 0 min: 100% B; 5 min: 10% A, 90% B; 10 min: 25% A, 60% B, 15% C; 13 min: 25% A, 60% B, 15% C; 17 min: 30% A, 10% B, 60% C; 18 min: 30% A, 10% B, 60% C; 21 min: 100%B.

^{15}N-THI5p overexpression and purification

The overexpression and purification procedures were the same as described above for THI5p except $^{14}NH_4Cl$ was replaced with $^{15}NH_4Cl$ (Cambridge Isotope Laboratories, Inc.) in the M9 salts.

Trypsin digestion of inactive THI5p

After completion of HMP-P formation, the mixture was desalted using a Bio-Spin 6 column (Bio-Rad) pre-equilibrated with 25 mM ammonium bicarbonate, pH 8.0. The inactive THI5p (50 μg, 1.8 μL of 660 μM THI5p) was added to 10 μL of 6 M guanidine-HCl, 25 mM ammonium bicarbonate, pH 8. 1 μL of 200 mM DTT in 25 mM ammonium bicarbonate pH 8.0, were then added to the denatured protein solution and the resulting mixture was incubated at room temperature for 1 hour. After incubation, 10 μL of 200 mM iodoacetamide in 25 mM ammonium bicarbonate pH 8.0, were added and the reaction mixture was incubated at room temperature in dark for 1 hour. Finally, 77.5 μL of 25 mM ammonium bicarbonate pH 8.0, were added to reduce the guanidine-HCl concentration to 0.6 M followed by the addition of 1 μg of trypsin. The mixture was incubated at 37 °C overnight and stored at -20 °C until analysis.

LC-MS conditions for analyzing THI5p samples after trypsin digestion

The following linear gradient was used at a flow rate of 0.4 mL/min over a Synergi Polar-RP 100A column (50 mm X 2 mm, 2.5 μm): solvent A is 0.1% formic acid in water, solvent B is 0.1% formic acid in acetonitrile; 0 min: 100% A; 1 min: 100%

A; 8 min: 88% A, 12% B; 31 min: 65% A, 35%B; 46 min: 35% A, 65% B; 48 min: 100% B; 53 min: 100%B; 56 min: 100%A; 65 min: 100% A. The HPLC is coupled with a microTOF-Q II for MS analysis in the positive mode.

The modified peptide tagged by aminooxy-biotin

The full reaction sample was dialyzed overnight against 10 mM HEPES, 5 M urea, pH 7.4. 50 µM of 1 mL of the sample was added with 100 µL of 30 mM aminooxy-biotin (Dojindo Lab) in water and 20 µL of 1.8 M NaOAc, pH 4.7. The mixture was incubated at room temperature for 24 h. 200 µg of the protein sample was digested with trypsin. After the digestion, the modified peptide with the aminooxy-biotin tag was purified by avidin resin. The purified peptide was eluted by the incubation of the avidin resin at 100 °C for 20 min. Finally, the sample was analyzed by LC-MS.

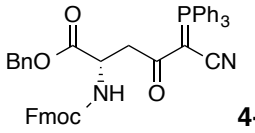
The modified peptide tagged by biotin hydrazide

The full reaction sample with 50 µg of protein was digested with trypsin. The mixture was added to an equal volume of 40 mM of biotin hydrazide in DMSO. The mixture was incubated at RT for 24 h and then dried by speedvac. The sample was then resuspended in 200 µL of water. Finally, the sample was analyzed by LC-MS.

The modified peptide tagged by methoxyamine

50 μg of protein from the full reaction sample were digested with trypsin. The mixture was added to an equal volume of 100 mM of methoxyamine in 0.2 M NaOAc, pH 5.4, and incubated at 37 °C for 24 h. Finally, the mixture was analyzed by LC-MS.

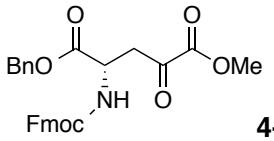
Synthesis of 4-6



4-6

2 g of Fmoc-Asp-OBz **4-4** (Bachem) and 1.35 g of (cyanomethylene)phosphorane **4-5** were added in 100 mL of CH_2Cl_2 under an argon atmosphere. The mixture was placed in an ice bath followed by the addition of 0.95 g of N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDCI) and 5.5 mg of 4-(Dimethylamino)pyridine (DMAP). Finally, the mixture was stirred at room temperature overnight. The solvent was removed after the reaction. The product was purified by silica gel (CH_2Cl_2/Ethyl acetate = 25/2). Finally, the product **4-6** is white solid with 75% yield (2.45 g). ^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.75 (d, 2H, *J* = 7.5 Hz, Ar*H*), 7.67~7.23 (26H, Ar*H*), 6.09 (d, 1H, *J* = 9 Hz, -N*H*-), 5.13 (dd, 2H, *J* = 12.6, 19.8 Hz, -CH_2-), 4.83-4.64 (m, 1H, -C*H*-), 4.51-4.37 (m, 1H, -C*H*-), 4.32-4.13 (m, 2H, -CH_2-), 3.66 (dd, 1H, *J* = 5.1, 16.8 Hz, -C*H*H-), 3.15 (dd, 1H, *J* = 4.2, 17.1 Hz, -CH*H*-). ^{13}C-NMR (75 MHz, $CDCl_3$): δ 193.3, 171.6, 156.2, 144.1, 143.8, 141.2, 133.6, 133.5, 133.3, 129.3, 128.5, 127.7, 127.6, 127.1, 125.3, 125.2, 123.2, 121.9, 119.9, 67.1, 66.9, 47.2, 40.8, 40.7. Observed $[M+H]^+$ = 729.2540 (Calculated $[M+H]^+$ = 729.2513).

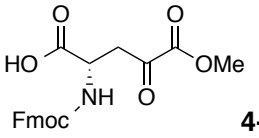
Synthesis of 4-7



4-7

2.45 g of **4-6** was dissolved in 100 mL of CH_2Cl_2/MeOH (7:3) under an argon atmosphere. The solution was brought to -78 °C followed by the purge of ozone. The reaction was stopped until the color of solution turned blue. Then the solution was purged with nitrogen to remove excess ozone in the solution at -78 ^{o}C. After the color of the solution disappeared, the solvent was removed. The product was purified by silica gel (CH_2Cl_2). Finally, the product **4-7** is white solid with 90% yield (1.48 g). ^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.75 (d, 2H, *J* = 7.5 Hz, Ar*H*), 7.56 (d, 2H, *J* = 7.5 Hz, Ar*H*), 7.43-7.27 (m, 9H, Ar*H*), 5.74 (d, 1H, *J* = 7.8 Hz, -N*H*-), 5.18 (dd, 2H, *J* = 12, 18.6 Hz, -CH_2-), 4.81-4.75 (m, 1H, -C*H*-), 4.44-4.31 (m, 2H, -CH_2-), 4.20 (t, 1H, *J* = 6.9 Hz, -C*H*-), 3.86 (s, 3H, -O*Me*), 3.51-3.47 (m, 2H, -CH_2-). ^{13}C-NMR (75 MHz, $CDCl_3$): δ 191.2, 170.1, 160.2, 155.8, 143.7, 143.6, 141.3, 134.8, 128.6, 128.5, 128.3, 127.7, 127.1, 125.1, 120.0, 67.8, 67.3, 53.2, 49.8, 47.0, 41.8. Observed $[M+Li]^+$ = 494.1770 (Calculated $[M+Li]^+$ = 494.1791).

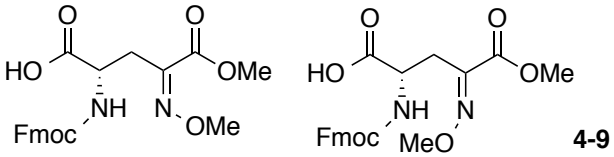
Synthesis of 4-8



4-8

1.48 g of **4-7** was dissolved in 40 mL of MeOH. Then 100 mg of 30% Pd/C was added into the solution. The reaction was performed under 1 atm of H_2 for 1 hour. After the reaction, Pd/C was filtered off by celite. Finally, the filtrate was dried and the product **4-8** was purified by silica gel. It is white solid with 85% yield (1.02 g). ^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.74 (d, 2H, *J* = 7.5 Hz, Ar*H*), 7.57 (d, 2H, *J* = 7.2 Hz Ar*H*), 7.41-7.26 (m, 4H, Ar*H*), 6.01-5.77 (*br*, 1H, -N*H*-), 4.81-4.73 (m, 1H, -C*H*-), 4.46-4.34 (m, 2H, -CH_2-), 4.20 (t, 1H, *J* = 6.9 Hz -C*H*-), 3.85 (s, 3H, -O*Me*), 3.67-3.19 (*br*, 2H, -CH_2-). ^{13}C-NMR (75 MHz, $CDCl_3$): δ 164.9, 156.1, 143.5, 141.2, 127.7, 127.1, 125.0, 120.0, 67.5, 46.9, 29.7. Observed $[M-H]^-$ = 396.1066 (Calculated $[M-H]^-$ = 396.1083).

Synthesis of 4-9



4-9

1.02 g of **4-8** and 0.32 g of methoxylamine hydrochloride were added in 20 mL of MeOH. The mixture was stirred at room temperature for 2 h. After the reaction, the solvent was removed. The solids were dissolved in CH_2Cl_2 followed by extraction with water three times. Finally, the CH_2Cl_2 portion was collected and dried by anhydrous sodium sulfate. After filtration and removing the solvent, the resulting product **4-9** was a

white solid (1.0 g, 95% yield). ^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.76 (d, 2H, *J* = 7.5 Hz, Ar*H*), 7.57 (m, 2H, Ar*H*), 7.40 (m, 2H, Ar*H*), 7.34-7.26 (m, 2H, Ar*H*), 5.63 (t, 1H, *J* = 8.7 Hz, -N*H*-), 4.69 (m, 1H, -C*H*-), 4.42-4.30 (m, 2H, -CH_2-), 4.22 (t, 1H, *J* = 6.9 Hz, -C*H*-), 4.08 (s, 3H, -O*Me*), 3.86 (s, 3H, -O*Me*), 3.75 (s, 3H, -O*Me*), 3.17-3.01 (m, 2H, -CH_2-). ^{13}C-NMR (75 MHz, $CDCl_3$): δ 174.6, 171.6, 164.0, 163.9, 156.0, 155.7, 147.1, 147.6, 143.8, 143.7, 143.6, 143.5, 141.2, 127.7, 127.0, 125.1, 125.0, 67.3, 67.2, 53.1, 53.0, 52.6, 51.2, 51.1, 47.0, 46.9, 29.7, 28.0, 27.6. Observed $[M-H]^-$ = 425.1328 (Calculated $[M-H]^-$ = 425.1349).

Solid phase peptide synthesis

The procedure followed the standard protocol. However, DMSO was added to a 20% final concentration in the amide coupling step from the installation of the modified residue **4-9**. Every amide coupling step was performed three times. After TFA cleavage, the peptide was dissolved in water and then analyzed by LC-MS. The data was analyzed by EIC and showed the desired peptide with methyl ester **4-10**. The peptide **4-10** solution was mixed with the same volume of 0.2 M LiOH in water. The mixture was incubated at room temperature for 1 h followed by LC-MS analysis. The result showed the desired peptide **4-11**.

PLP by-product trapped by phenylhydrazine

After the reaction finished, the mixture was filtered using a 10 kDa cut-off filter. 10 μL of phenylhydrazine stock solution (5 μL in 1 mL of water) was added to the filtrate. The mixture was incubated at 37 °C for 2.5 h followed by HPLC analysis.

HPLC conditions for analyzing the phenylhydrazine derivatizing reaction

The following linear gradient, at a flow rate of 2 mL/min, on a Supelcosil LC-18 column (250 mm X 10 mm, 5 μm ID) was used: solvent A is water, solvent B is 100 mM KH_2PO_4, pH 6.6, solvent C is methanol; 0 min: 100% B; 4 min: 25% A, 60% B, 15% C; 8 min: 20% A, 40% B, 40% C; 14 min: 10% A, 10% B, 80% C; 18 min: 10% A, 10% B, 80% C; 20 min: 25% A, 50% B, 25% C; 22 min: 100%B; 30 min: 100% B.

HPLC conditions for collecting the phenylhydrazine derivatizing reaction

The condition is identical to the previous one. But solvent B is 100 mM NH_4OAc, pH 6.6.

LC-MS conditions

The following linear gradient, at a flow rate of 0.4 mL/min, on a Supelcosil LC-18-T column (150 mm X 3 mm, 3 μm ID) was used: solvent A is 20 mM NH_4OAc, pH 6.6, solvent B is 75% methanol; 0 min: 100% A; 3 min: 100% A; 8 min: 50% A, 50% B; 10 min: 100% A. The HPLC was coupled with a microTOF-Q II for MS analysis in the positive mode.

Yeast transformation

INV*Sc*1 *S. cerevisiae* was inoculated into 30 mL of YPD with shaking overnight at 30^{o}C. Then, 2 mL of culture was transferred to 30 mL of YPD to make a final OD_{600} of 0.1. The culture was incubated with shaking at 30^{o}C until OD_{600} reached 0.5 (at least two generations) followed by centrifugation at 3000 g for 5 min. The supernatant was discarded and 20 mL of water was added to resuspend the cells. The mixture was again centrifuged at 3000 g for 5 min and the supernatant was discarded. Finally, 500 µL of water was added to resuspend the cells. Concurrently, 360 µL of transformation mix was prepared. The composition of the transformation mix is 240 µL of 50% W/V of PEG3550, 36 µL of 1 M of LiOAc, 50 µL of boiled single strained carrier DNA (2 mg/mL) (Salmon sperm DNA, Sigma), and 34 µL of YesNTA2 plasmid solution containing the *ScTHI5* gene in water. 100 µL of resuspended cell solution was mixed with 360 µL of the transformation mix. The mixture was vigorously vortexed and incubated in a 42^{o}C water bath for 40 min. Then the mixture was centrifuged at 14000 rpm for 30 sec and the supernatant was removed. Finally, 1 mL of water was added to resuspend the cells, and the cell solution was plated on SC-Ura selection medium and incubated at 30 °C for two to three days.

Overexpression and purification of ScTHI5p in INVSc1 S. cerevisiae

After successful transformation, the colony was inoculated in 30 mL of Sc-Ura containing 1% raffinose and 2% galactose and grown until turbid. Then the culture was

added into 1.5 L of the same media and grown an additional 3 days. Cells were harvested by centrifugation and stored at -80 °C until purification.

12 g of cell from 5 L of culture was thawed and resuspended in lysis buffer with 2 tablets of protease inhibitor cocktail (cOmplete EDTA free, Roche). The cells were lysed using a microfluidizer processor (M110P, Microfluidics) with 28000 psi three times. The resulting cell lysate was clarified by centrifugation and the THI5 protein was purified on a Ni-NTA column following the manufacturer's instructions. The purified THI5p solution was buffer-exchanged to a desalting buffer (100 mM Tris, 30% glycerol, pH 7.5) and stored at -80 °C.

Results and discussion

Identification of the modified His66 of inactive THI5p

In the previous chapter, His66 of THI5p was identified as N=C-N donor for HMP-P formation. However, there was no literature report about the structure of histidine after HMP-P formation. Furthermore, this reaction is unprecedented in terms of chemistry and biochemistry, making it important to find an approach to identify it experimentally.

The strategy for this identification was to digest inactive THI5p followed by LC-MS and MS/MS analysis (Figure 4-2). The candidate peptide containing the modified His66 was residues 63-71. Because there was no proposed structure for the modified His66, extracted ion chromatography (EIC) could not be used to identify the desired

peptide. However, the MS/MS analysis of residues 63-71 of the native peptide (the sequence is AMVHTLAAK) gave a very helpful hint. The signals of a_2, b_2 and y_2 ions of this peptide were always good. In addition, these three ions corresponded to AM at the N-terminus and AK at the C-terminus. In the predicted peptide lists of THI5p by trypsin, only peptide 63-71 has these three ions present at the same time. Hence, the EIC was used in the MS/MS analysis to identify the appearance of these three ions at the same retention time. Finally, a peptide which $[M+2H]^{2+}$ is 474.2 Da was selected for further investigation.

This peptide was 5.96 Da larger than the native peptide 63-71. In order to solve the peptide sequence, we assumed +5.96 Da at His66 to assign the corresponding ions in the MS/MS spectrum. Most of ions could be assigned to confirm the peptide sequence with modified His66 (labeled as X in Figure 4-2). The most important information was the y_5 ion of this peptide corresponding to the fragment from the C-terminus to Thr67 was identical to y_5 ion of the native peptide 63-71. The y_6 ion of this peptide was 6 Da larger than the y_5 ion of the native peptide 63-71.

In order to analyze the MS/MS spectrum correctly, inactive ^{15}N-THI5p was applied with the same analysis. In the MS/MS spectrum, all of ions had molecular weight (MW) shifts due to ^{15}N incorporations. Moreover, the most informative thing is that the MW difference of the inactive ^{15}N-THI5p between y_5 and y_6 was 144 Da compared with 143 Da in inactive ^{14}N-THI5p, suggesting that there is one nitrogen atom at the modified His66. This 1 Da shift was due to the ^{15}N of the amide. His66 contains

three nitrogen atoms, so these results indicate that the modified His66 does not contain the two nitrogen atoms on the imidazole ring.

All of analysis indicated that His66 increased by 6 Da after HMP-P formation. Furthermore, it was also the first direct evidence to show His66 is the N=C-N donor.

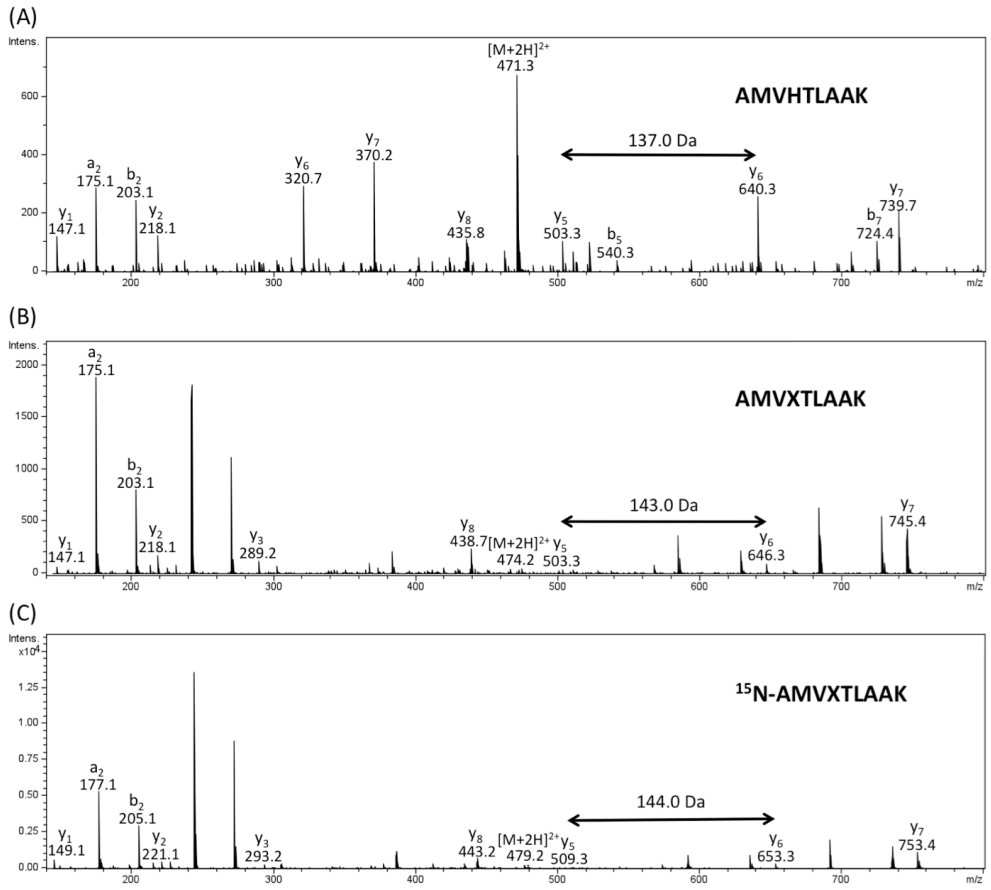


Figure 4-2. The MS/MS spectra of peptides for peptide 63-71 (A), modified peptide 63-71 (B), and ^{15}N-modified peptide 63-71 (C).

Characterization of the modified His66

The results indicated that His66 gains 5.96 Da after HMP-P formation, but the lack of structural information for the modified His66 (X), meant that the only approach was to test the functional groups on the modified His66. First I tested for the presence of ketone or aldehyde group which could be reduced by $NaBH_4$. This would result in a 2 Da increase after $NaBH_4$ treatment. Therefore, the pool of the peptides derived from trypsin digestion of the inactive THI5p was treated with excess of $NaBH_4$ followed by LC-MS analysis. The EIC was able to identify a peptide with a 2 Da increase and the MS-MS analysis confirmed the peptide sequence with 7.96 Da increase at His66 (Figure 4-3A). The MW difference between the y_5 and y_6 ions were 145 Da compared with the previous spectrum. It indicated that the MW increase was from the modified His66 implying that the modified His66 has a ketone or an aldehyde group.

Because of this, the aminooxy-biotin tag, which is used to study glycoproteins by the formation of oxime, was also used to enrich the peptide containing the modified His66 (Figure 4-3 B) [21]. The pool of the peptides derived from the trypsin digestion of the inactive THI5p was incubated with the tag followed by avidin resin enrichment. LC-MS analysis was used to detect the resulting peptide, however, the peptide signal was insufficient to get the MS/MS spectrum. But when the same procedure was applied on the inactive ^{15}N-THI5p, LC-MS analysis was able to detect the ^{15}N-peptide with 10 Da shift, implying that 10 nitrogen atoms remained on that peptide after His66 donated two nitrogen atoms from the imidazole ring for HMP-P synthesis.

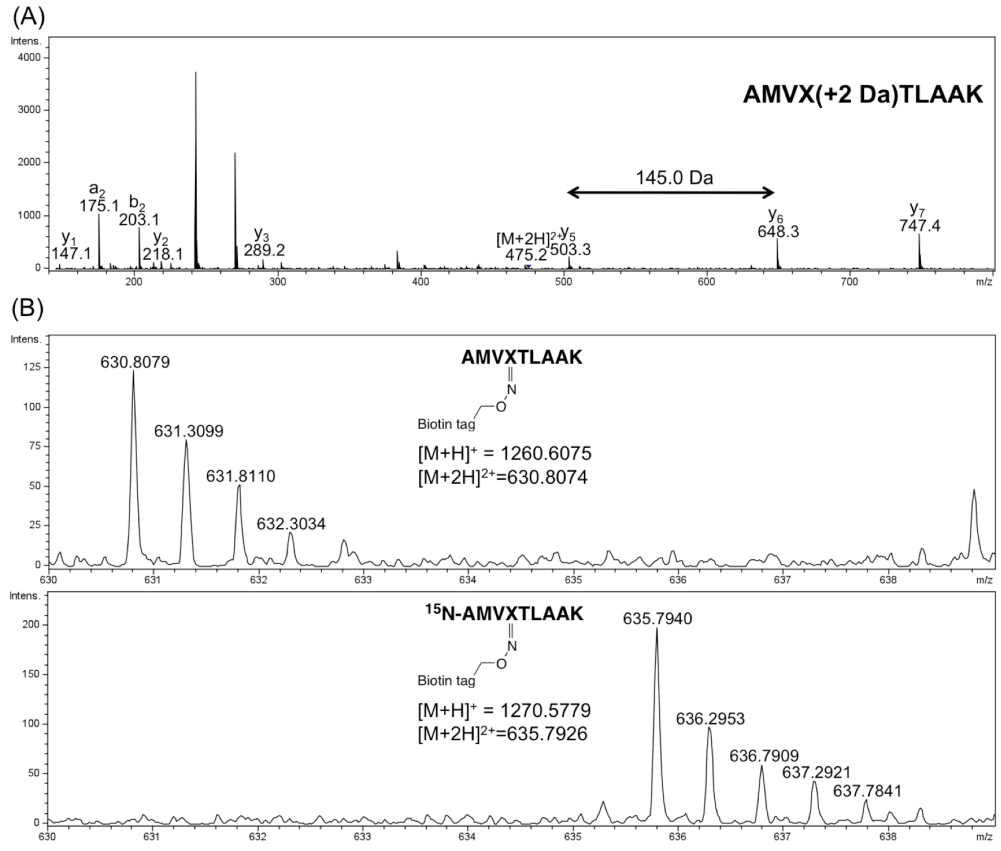


Figure 4-3. (A) The MS/MS spectrum of modified peptide with $NaBH_4$ treatment showing X with +2 Da. (B) The MS spectra of ^{14}N- and ^{15}N-modified peptide with the aminoxy-biotin tag.

During the aminoxy-biotin tag investigation, biotin hydrazide was used to enrich the modified peptide by the formation of hydrazone and EIC was used to analyze the corresponding peptide. But it was the MS/MS spectrum that proved the most important and interesting (Figure 4-4A). The spectrum confirmed the peptide sequence and that

there were a few ions showing -44 Da for the predicted ions. This implied that the hydrazone was involved in decarboxylation during CID fragmentation.

The results led to the postulation that the structure of the modified histidine is a keto-acid. This proposed structure corresponds with a 5.96 Da increase and a ketone group. Concerning the decarboxylation in the MS/MS fragmentation, biotin hydrazide forms hydrazone followed by tautomerization (Figure 4-4B), which could readily undergo decarboxylation. This may be applicable to proteomic studies of biological compounds containing keto-acids.

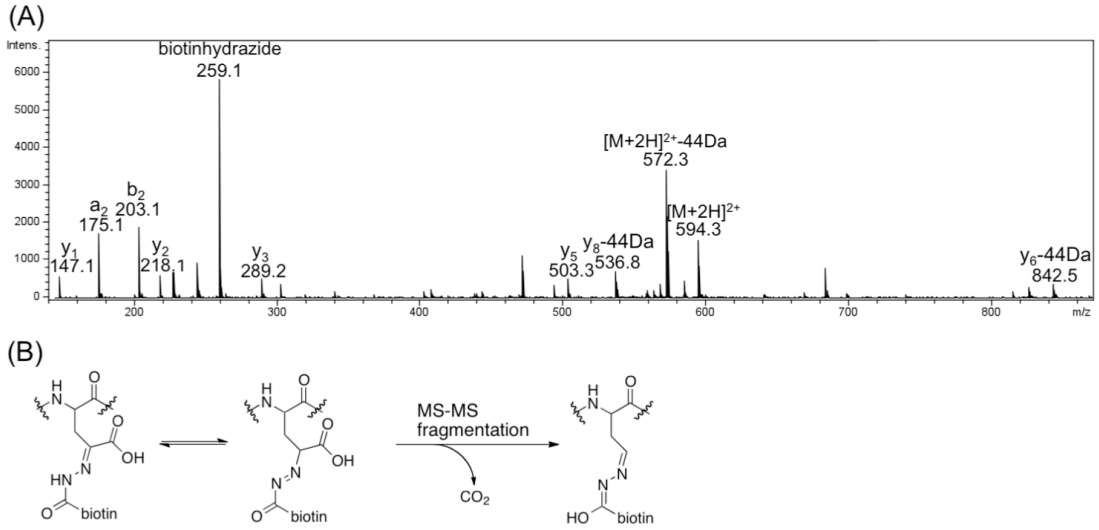


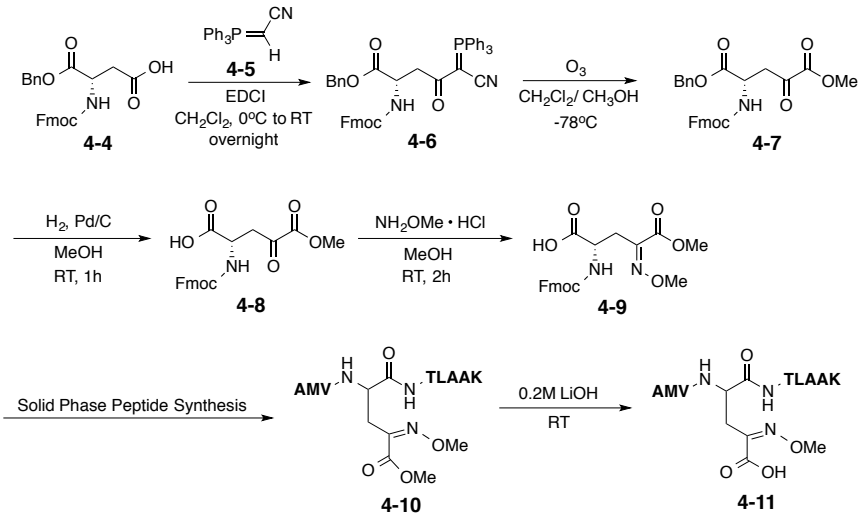
Figure 4-4. (A) The MS/MS spectrum of modified peptide with biotinhydrazide showing the decarboxylation fragment. (B) The proposed mechanism of decarboxylation suggests a keto-acid structure.

Synthesis of the reference peptide

By using $NaBH_4$ and biotin hydrazide, the structure of His66 after HMP-P synthesis was proposed as its imidazole becomes a keto-acid. Although the peptide analysis were quite convincing, this enzymatic reaction is unprecedented. I decided to synthesize the reference peptide to compare it to the modified peptide.

The synthetic route is shown in Scheme 4-1. Protected glutamate **4-4** was reacted with (cyanomethylene)triphenylphosphorane **4-5** in the presence of EDCI to form cyano-keto phosphorane **4-6** [22]. It underwent oxidative cleavage with ozone in the presence of methanol to form an α-keto ester product **4-7**. The α-keto ester product **4-7** was deprotected by hydrogenolysis to generate **4-8**, which was then protected by methoxyamine to conduct solid phase peptide synthesis [23,24]. Finally, the ester of the α-keto ester of the peptide **4-10** was hydrolyzed under a basic condition to generate the reference peptide **4-11**.

Since the reference peptide had oxime protection by methoxyamine, the pool of peptide containing the modified-His66 peptide was incubated with an excess of methoxyamine. Finally, the resulting peptide could be co-migrated with the reference peptide by LC-MS and MS/MS analysis. The EIC analysis showed that these two peptides co-migrated well. Most importantly, the MS/MS analysis of these two peptides showed identical fragmentation. These results confirm that the imidazole of His66 becomes a keto-acid after HMP-P synthesis.



Scheme 4-1. The synthetic route of reference peptide **4-11**.

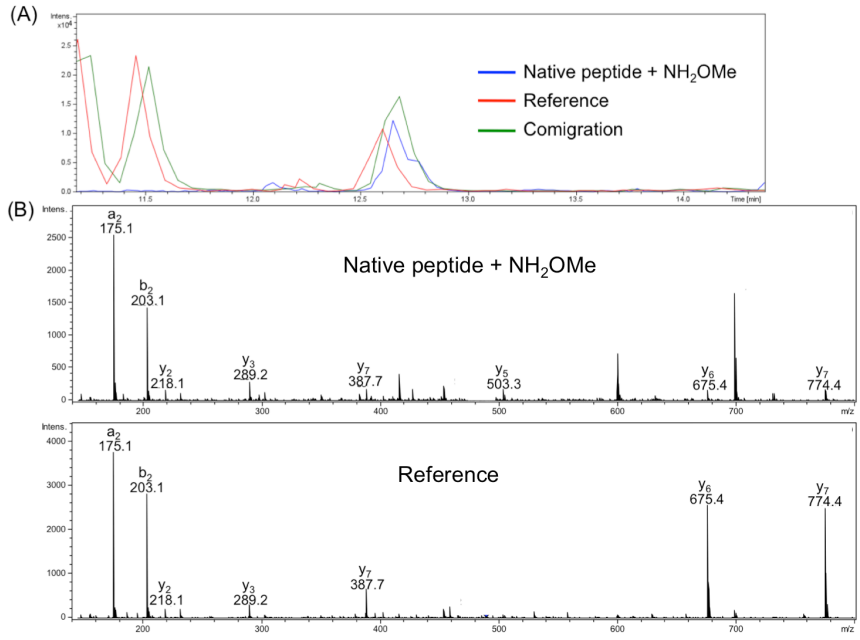


Figure 4-5. (A) The EIC analysis of co-migration. (B) The MS/MS spectra of the native peptide with methoxyamine and the reference peptide showed identical fragmentation.

Oxygen labeling of the His66-derived keto-acid

The THI5p assay requires O_2 to generate HMP-P. Because the His66-derived keto-acid contains oxygen atoms, we wanted to identify the source of the oxygen atoms. The full reaction assay was incubated under an atmosphere of $^{18}O_2$ followed by peptide analysis (Figure 4-6).

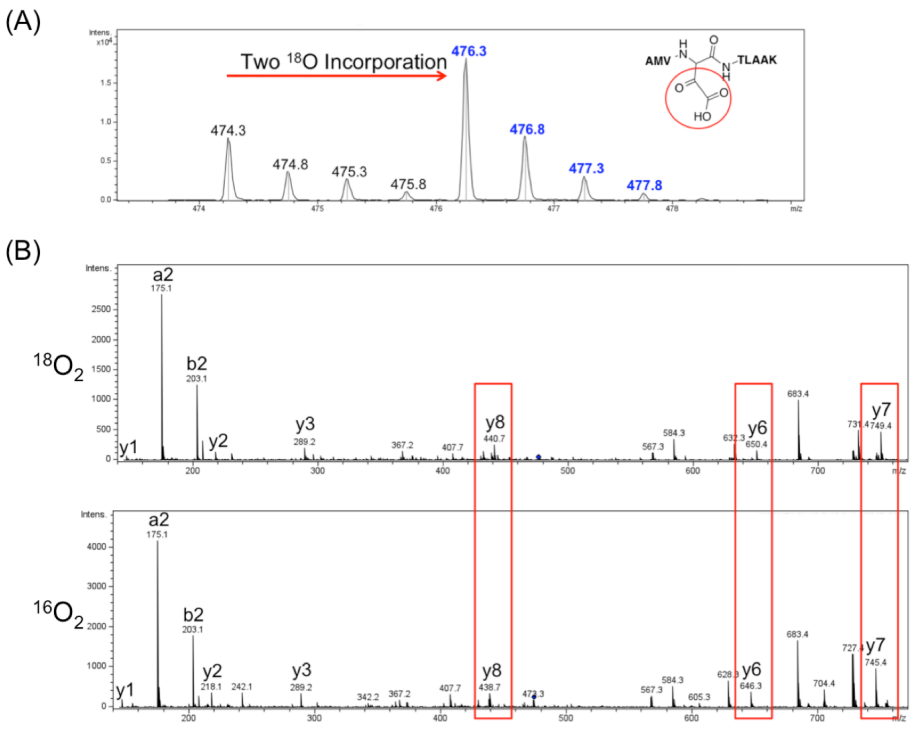


Figure 4-6. (A) The MS spectrum of the modified peptide showing two ^{18}O incorporations. (B) The MS/MS spectra of the modified peptide derived from $^{16}O_2$ and $^{18}O_2$ atmosphere to identify ^{18}O incorporation at the His66-derived keto-acid.

Surprisingly, the MW of the modified peptide had +4 Da corresponding to two ^{18}O incorporations. Notably, the spectrum included non-labeled peptide because THI5p underwent the full reaction during overexpression. This did not hinder the MS/MS spectra which showed that the y_6 ion had a 4 Da increase in the ^{18}O-labeled peptide indicating ^{18}O incorporation at the modified His66.

However, the keto-acid structure has three oxygen with two ^{18}O incorporations. There are two possible patterns for ^{18}O incorporation: two ^{18}O at the carboxylic acid or one ^{18}O at the ketone and one ^{18}O at the carboxylic acid. To determine where incorporation was occurring we used methoxyamine to form an oxime in the ^{18}O-labeled modified peptide (Figure 4-7). The resulting peptide still had +4 Da indicating incorporation of the two ^{18}O at the carboxylic acid.

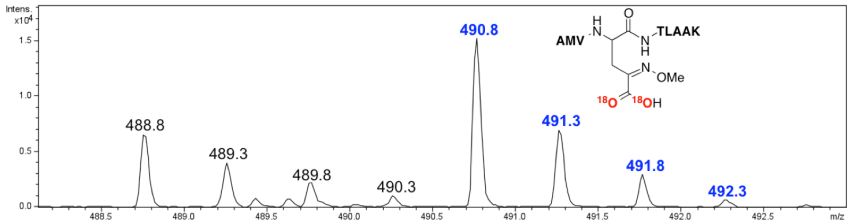


Figure 4-7. The ^{18}O incorporating peptide with methoxyamine treatment to identify ^{18}O incorporating positions at the carboxylic acid.

Since there are two ^{18}O incorporating at the carboxylic acid of the keto-acid, it is interesting to know whether these two ^{18}O were derived from the same or different O_2. The assay was incubated under an atmosphere composed of a mixture of $^{16}O_2$ and $^{18}O_2$

and the MW of the modified peptide was analyzed (Figure 4-8). Interestingly, the isotope distribution of the MW of the modified peptide indicated the presence of $^{16}O/^{16}O$, $^{16}O/^{18}O$, and $^{18}O/^{18}O$ labeling meaning that the two oxygen atoms in the carboxylic acid are from two different O_2.

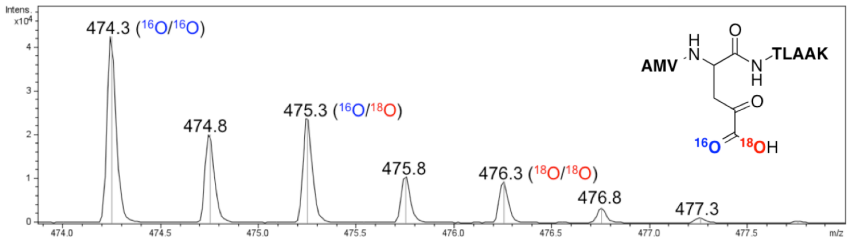


Figure 4-8. The MS spectrum of modified peptide derived from the mixture of $^{16}O_2$ and $^{18}O_2$ atmosphere.

In vivo *investigation of the presence of keto-acid in THI5p*

The characterization of the His66-derived keto-acid was accomplished *in vitro*. But is the His66-derived keto-acid formed *in vivo*? To investigate this *Sc*THI5p was overexpressed in both *S. cerevisiae* and *S. cerevisiae* Δ*THI5*. Each was grown in minimal media without thiamin supplementation. The overexpressed THI5p was purified by Ni-NTA chromatography, digested by trypsin, and analyzed by LC-MS. The candidate peptide sequence (AM**I**HTLAAK) was different from *Ca*THI5p (AM**V**HTLAAK), but the EIC analysis can identify the desired peptide using the

predicted MW. The MS/MS analysis confirmed the peptide sequence and the y_6 ion showed a 6 Da increase compared with the original peptide (Figure 4-9).

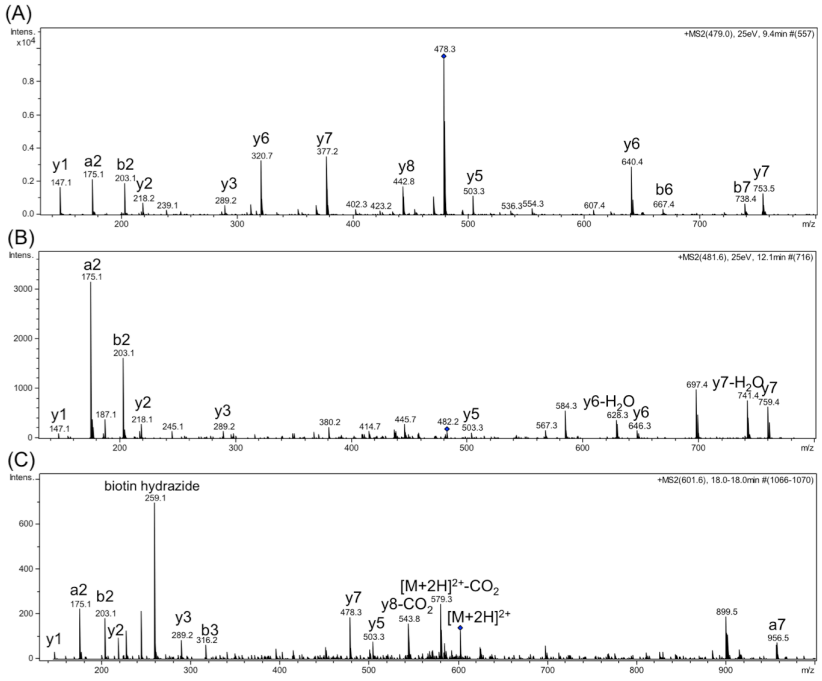


Figure 4-9. The MS/MS spectra of the *S. cerevisiae* original peptide (A), modified peptide (B), and modified peptide with biotin hydrazide (C).

All of experiments were repeated with the same result. These experiments indirectly demonstrated the His66-derived keto-acid forms after HMP-P formation.

Identification of the PLP by-product

The feeding studies demonstrated that PLP partially incorporates into HMP-P. The question is what PLP by-product forms after HMP-P formation? PLP has an aldehyde group that forms an imine with Lys62 and this aldehyde does not incorporate into HMP-P. We tested the hypothesis that the PLP by-product might retain this aldehyde group. Normally, 2,4-dinitrophenylhydrazine is normally used to trap aldehyde with the formation of hydrazone, which has a unique λ_{max} between 300 to 400 nm. This specific λ_{max} monitored by HPLC analysis could readily identify trapping compounds. However, 2,4-dinitrophenyl hydrazine does not dissolve very well in an aqueous buffer, and can damage the reverse phase column. Furthermore, the assay conditions are acidic which might degrade acid sensitive substrates. Therefore, phenylhydrazine was chosen for this investigation because of its good solubility in aqueous solution and neutral assay conditions.

After completion of the reaction, the protein was removed followed by the addition of phenylhydrazine solution. Finally the mixture was analyzed by HPLC monitored at 350 nm. The HPLC analysis showed one peak in the THI5p wild-type full reaction instead of other three control samples (Figure 4-10).

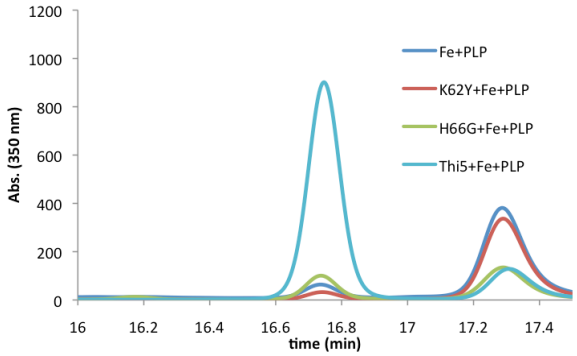


Figure 4-10. The HPLC analysis of phenylhydrazine trapping of the PLP by-product.

Moreover, the THI5p H66G and K62Y proteins did not have this peak in HPLC analysis. This peak was collected and analyzed by LC-MS (Figure 4-11). Surprisingly, the collected pool contained two compounds, which were separated by the small particle size column used for LC-MS. The first peak was identified by MS and MS/MS as a glyoxylate and phenylhydrazine adduct; the reference compound was synthesized to conduct a co-migration experiment. The second peak based on the MS and MS/MS analysis was proposed to be a C3-fragment and phenylhydrazine adduct. In this case the reference compound was not made because of the difficulty of the synthesis. However, the MS/MS analysis gave enough information to assign the adduct structure.

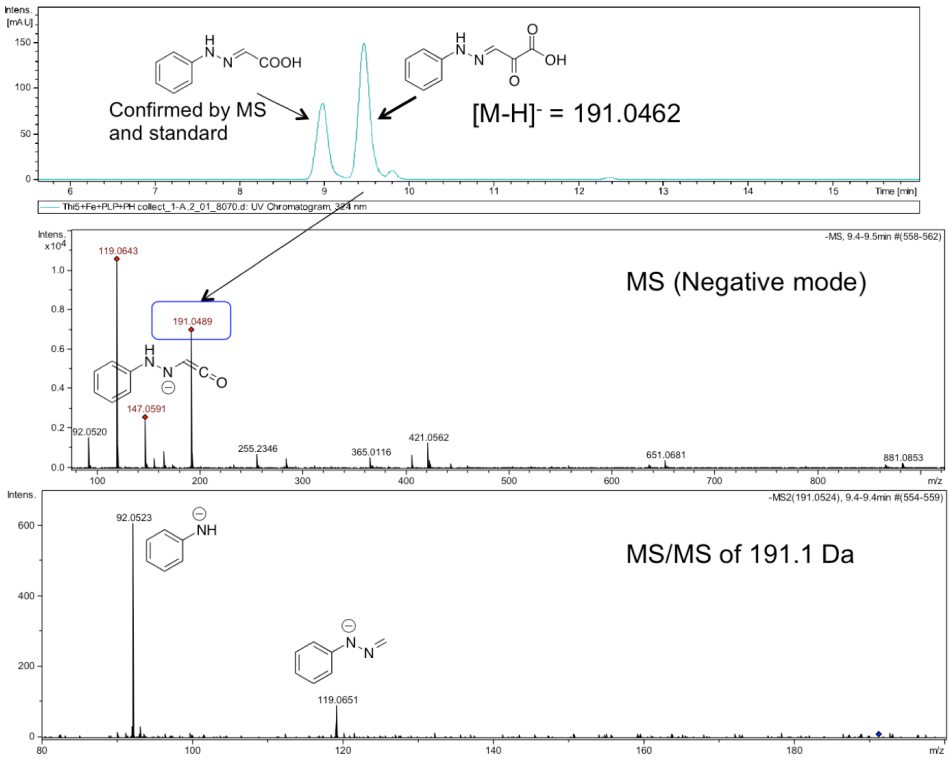


Figure 4-11. The LC-MS and MS/MS analysis of the PLP by-product identified as glyoxylate and a C3-fragment.

To confirm these two molecules were derived from the THI5p catalyzed reaction, 4',5'-^{13}C-PLP was used to set up the full reaction assay. The MW of HMP-P had +1 Da as the prediction to suggest the reaction works well. In the meantime, the assay after the removal of THI5p was derivatized by phenylhydrazine. Finally, the pool containing these two adducts was collected and analyzed by LC-MS (Figure 4-12). The MW of the C3-fragment adduct showed +1 Da. Likewise, the MS of the glyoxylate adduct also had a 1 Da increase based on isotopic distribution. The reason the glyoxylate adduct did not

have a complete $[M+1]^+$ signal is because the glyoxylate was present in both Eppendorf tubes and glycerol. Therefore, the assay using 4',5'-^{13}C-PLP could further confirm that these two molecules derived from PLP. Moreover, the MS/MS analysis of these two adducts could identify the position of ^{13}C. In both of adducts, the ^{13}C was at the imine carbon, suggesting that the aldehyde group of PLP is retained after HMP-P formation.

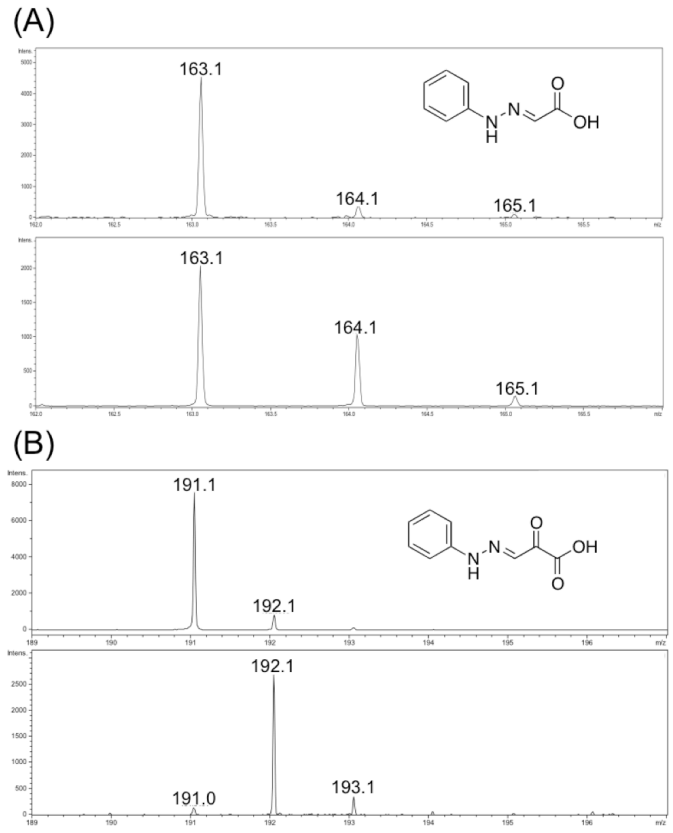


Figure 4-12. The MW analysis of glyoxylate (A) and the C3-fragment (B) with the phenylhydrazine derivative showing one ^{13}C incorporation from 4',5'-^{13}C-PLP.

Decarboxylation of the C3-fragment to generate glyoxylate

In the phenylhydrazine trapping experiments, two molecules were identified as PLP by-products, but the final product was proposed to be glyoxylate. To compare these two molecules, it was proposed that the C3-fragment containing keto-acid undergoes decarboxylation to generate glyoxylate. In the literature, the decarboxylation of keto-acid could undergo radical reaction, metal-assisted reaction involving radical, and H_2O_2-catalyzed reaction [25,26].

During the course of the THI5p studies, other biochemical investigations suggested that H_2O_2 catalyzed the decarboxylation of the keto-acid. A pool of these two molecules and phenylhydrazine adducts was collected, treated with H_2O_2, and analyzed by HPLC (Figure 4-13). The HPLC analysis comparing the samples before and after H_2O_2 treatment showed that the glyoxylate adduct increased concomitant with the disappearance of the C3-fragment adduct. This suggested that H_2O_2 reacted with the C3-fragment to form glyoxylate following decarboxylation.

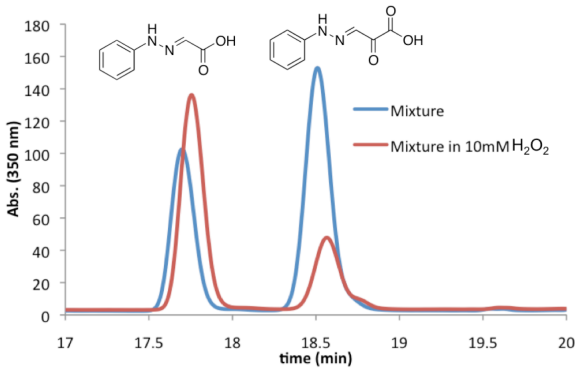


Figure 4-13. The HPLC analysis of the decarboxylation of the keto-acid catalyzed by H_2O_2.

Oxygen labeling of the C3-fragment and glyoxylate

Since the His66-derived keto-acid has two oxygen atoms incorporated by O_2, we investigated oxygen labeling using the same strategy as for the THI5p peptide.

The full reaction assay was incubated under an atmosphere of $^{18}O_2$. After the reaction was complete, the protein was removed followed by phenylhydrazine derivatization. The pool containing these two adducts was collected and analyzed by LC-MS (Figure 4-14).

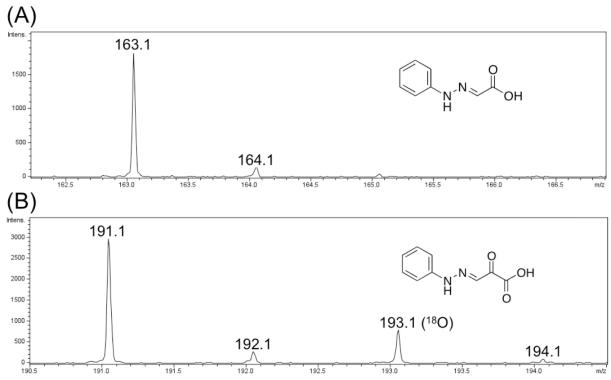


Figure 4-14. The MS spectra of glyoxylate (A) and the C3-fragment (B) derived from the $^{18}O_2$ atmosphere.

The MW of the C3-fragment adduct increased by 2 Da corresponding to one oxygen incorporation. The MS/MS analysis identified the ^{18}O position at the carboxylic acid. In addition, the MW of the glyoxylate adduct showed 2 and 4 Da increase. The partial incorporation suggested that solvent exchange had occurred. Therefore, ^{18}O-water

was used to investigate the exchangeable group. The mixture of these two adducts was lyophilized completely and redissolved in ^{18}O-water. The glyoxylate adduct did not show any incorporation of ^{18}O, but the C3-fragment adduct had one ^{18}O incorporation and the MS/MS analysis identified that it incorporated at the ketone. Moreover, the results also indicated that the carboxylic acid does not undergo solvent exchange.

The full reaction assay was set up in the buffer containing 60% ^{18}O-water. After phenylhydrazine derivatization, the mixture was analyzed by LC-MS (Figure 4-15). The MW of the C3-fragment adduct showed about 60% ^{18}O incorporation. The MW of the glyoxylate adduct showed the incorporation of only one ^{18}O. Thus, the solvent exchange happened before phenylhydrazine derivatization. But it explains why the MW of the glyoxylate adduct in the $^{18}O_2$ assay showed both one and two ^{18}O incorporations.

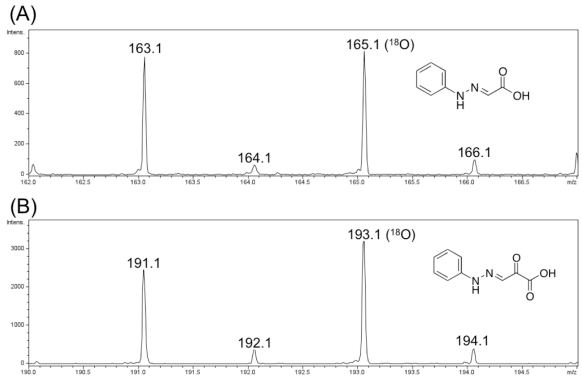
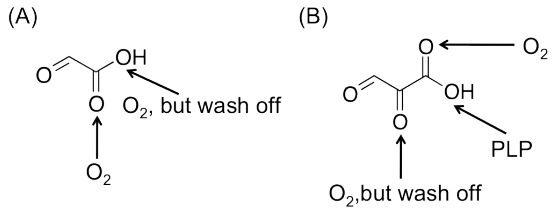


Figure 4-15. The MS spectra of glyoxylate (A) and the C3-fragment (B) derived from 60% ^{18}O-water conditions.

To summarize these labeling results, the positions of oxygen incorporation are shown in Scheme 4-2.



Scheme 4-2. The oxygen labeling pattern of glyoxylate (A) and the C3-fragment (B).

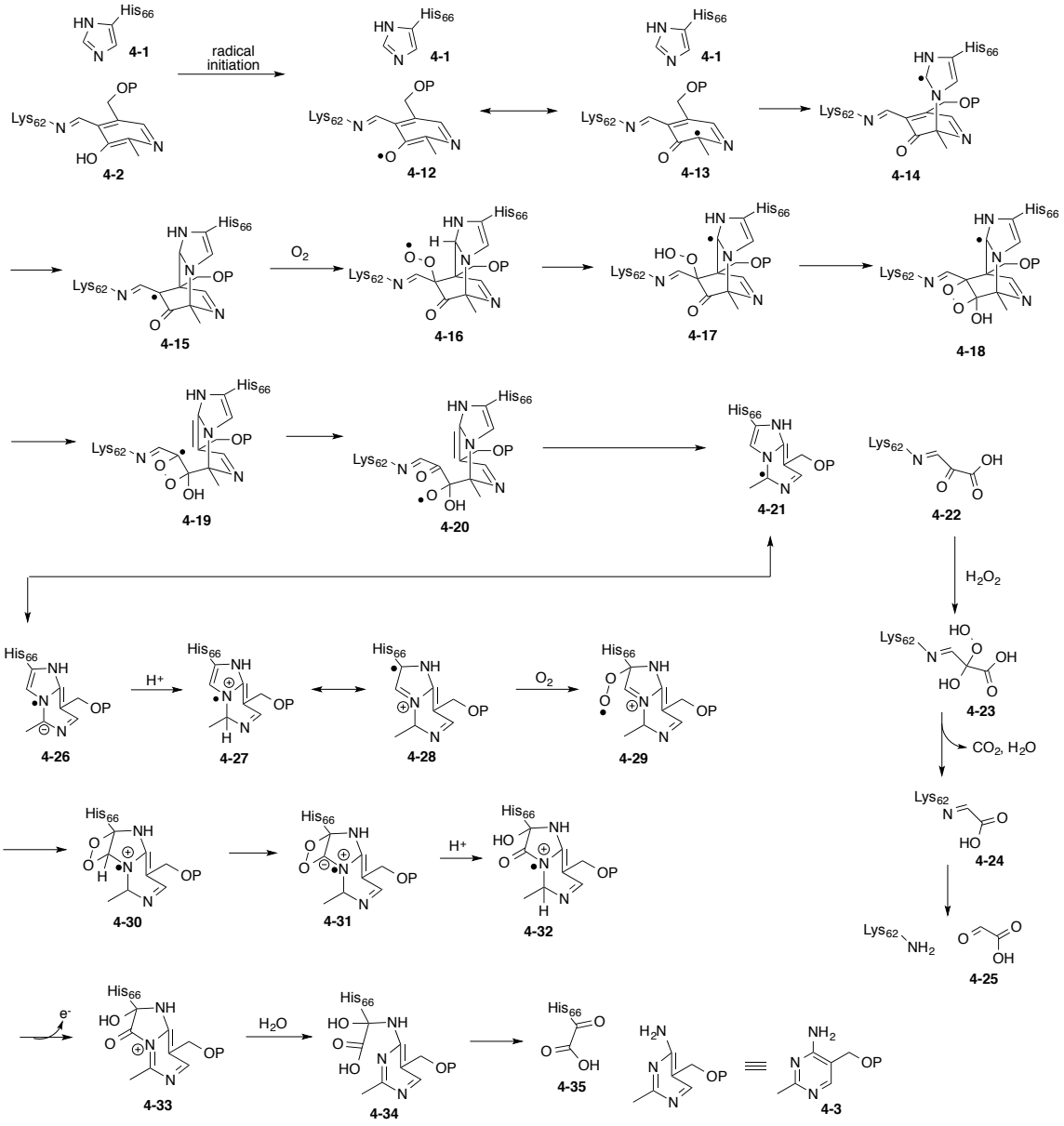
Conclusion

In this chapter, the His66 structure and PLP by-products after HMP-P formation were characterized. The imidazole of His66 becomes a keto-acid and the PLP by-product goes through the C3-fragment as an intermediate to generate glyoxylate after decarboxylation. Oxygen incorporates into both the keto-acid and glyoxylate. This will help us complete the picture of the mechanism of HMP-P formation (Scheme 4-3).

The reaction is initiated by the formation of a phenolic radical **4-12**. The radical adds to the imidazole ring of His66 to form **4-14** followed by the addition to C5 of PLP to generate **4-15**. **4-15** reacts with O_2 to form a superoxide radical **4-16** followed by hydrogen abstraction to generate **4-17**. The peroxy group of intermediate **4-17** attacks the ketone to form **4-18**. The intermediate **4-18** undergoes a series of β-scission reactions to generate **4-21** and **4-22**. The keto-acid intermediate **4-22** reacts with hydrogen peroxide to undergo decarboxylation which generates glyoxylate **4-25**. Hydrogen

peroxide is proposed to be generated from oxygen reacts with di-ferrous (the investigation of di-ferrous on THI5p is in Chapter 7). The structure of **4-22** is similar to the C3-fragment with phenylhydrazine. The investigation of the PLP by-product could support this part of proposed mechanism.

4-21 becomes **4-28** after resonance and tautomerization. The intermediate **4-28** reacts with another O_2 to generate **4-29**. The superoxide radical of **4-29** attacks iminium to form **4-30** followed by deprotonation to break the O-O bond to generate **4-32**. **4-32** undergoes $1e^-$ oxidation and deprotonation to form **4-33** followed by an attack of water to generate **4-34**. The water is proposed to be the by-product from O_2 catalyzed by iron. Finally, **4-34** undergoes elimination to generate the His66-derived keto-acid **4-35** and HMP-P **4-3**. This proposed mechanism matches the labeling results in the $^{18}O_2$ experiments.



Scheme 4-3. The proposed mechanism of HMP-P formation catalyzed by THI5p.

CHAPTER V

INVESTIGATION OF THE PLP-DERIVED SHUNT REACTION AND MODIFICATIONS ON THI5 PROTEIN

Introduction

In the last two chapters, the reconstitution of THI5p activity and the characterization of the modified His66 and PLP by-products after HMP-P formation were explored. Moreover, oxygen is also incorporated into the His66-derived keto-acid and PLP by-products. In this chapter, a PLP shunt product was identified to support a radical initiation mechanism. Cys199 and Met320 were oxidized during HMP-P formation. Lastly, 2-oxo-histidine was characterized as a by-product.

Experimental methods

PLP-nitrile identification

100 μL of 600 μM THI5p was anaerobically pre-incubated with 2 molar equivalents of $Fe(NH_4)_2(SO_4)_2$ in an ice bath for 30 min. The mixture was anaerobically desalted using a Bio-Spin 6 column (Bio-Rad) pre-equilibrated with 100 mM Tris-HCl, pH 7.5, followed by the addition of 2 molar equivalents of PLP. The reaction mixture was aerobically incubated at room temperature for 3 h, filtered using a 10 kDa MW cut-off filter, and analyzed by HPLC and LC-MS.

HPLC conditions for analyzing and isolating PLP-nitrile

The following linear gradient was used, at a flow rate of 2 mL/min, on a Supelcosil LC-18 column (250 mm X 10 mm, 5 μm ID): solvent A is water, solvent B is 100 mM KH_2PO_4, pH 6.6, solvent C is methanol; 0 min: 100% B; 5 min: 10% A, 90% B; 10 min: 25% A, 60% B, 15% C; 13 min: 25% A, 60% B, 15% C; 17 min: 30% A, 10% B, 60% C; 18 min: 30% A, 10% B, 60% C; 21 min: 100% B. For the isolation, solvent B is 100 mM NH_4OAc, pH 6.6.

Preparation of THI5p hydrolysate for 2-oxo-histidine investigation

100 μL of 600 μM THI5p was anaerobically pre-incubated with 2 molar equivalents of $Fe(NH_4)_2(SO_4)_2$ in an ice bath for 30 min. The mixture was anaerobically desalted using a Bio-Spin 6 column (Bio-Rad) pre-equilibrated with 100 mM Tris-HCl, pH 7.5. The reaction mixture was aerobically incubated at room temperature for 3 h. Then the mixture was buffer exchanged to 20 mM ammonium bicarbonate using a Bio-Spin 6 column. The resulting solution was dried by lyophilization and redissolved in 2 mL of 6 M HCl and 40 mM DTT. The mixture was refluxed at 105 °C for 24 h and dried by lyophilization. The resulting solid was redissolved in water containing 0.1% acetic acid for further investigation.

HPLC conditions for isolating the fraction containing 2-oxo-histidine

The following linear gradient, at a flow rate of 1 mL/min, on a Supelcosil LC-18-T column (150 mm X 4.6 mm, 5 μm ID) was used: solvent A is water containing 0.1%

acetic acid, solvent B is acetonitrile containing 0.1% acetic acid; 0 min: 100% A; 5 min: 100% A; 10 min: 100% B; 13 min: 100% B; 16 min, 100% A; 25 min, 100% A.

Ortho-*phthaldialdehyde (OPA) derivatization of amino acid analysis*

25 μL of THI5p hydrolysate or 2-oxo-hisitinde solutions was added into 30 μL of 20 mM OPA, 1 M potassium borate, pH 8.0, 60 mM β-mercaptoethanol, 1.5% methanol. The mixture was incubated at room temperature for 1 min and then analyzed by HPLC.

HPLC conditions for analyzing OPA derivatization of THI5p hydrolysate

The following linear gradient, at a flow rate of 2 mL/min, on a Supelcosil LC-18 column (250 mm X 10 mm, 5 μm ID) was used: solvent A is 50 mM NaCl, solvent B is methanol; 0 min: 100% A; 16 min: 75% A, 25% B; 22 min: 60% A, 40% B; 25 min, 50% A, 50% B; 31 min, 45% A, 55% B; 35 min, 25% A, 75% B; 40 min, 25% A, 75% B; 43 min, 100% A; 53 min, 100% A.

Results and discussion

Identification of PLP-derived shunt product

During investigating the reconstitution of THI5p, HPLC analysis monitored at 230 nm occasionally showed a peak near the void volume and the UV spectrum of this peak was similar to PLP (Figure 5-1). After careful investigation, this peak reached a maximum when the assay did not contain DTT. Moreover, this peak was derived from

an enzymatic reaction. Heat-denatured THI5p cannot generate this under the same conditions.

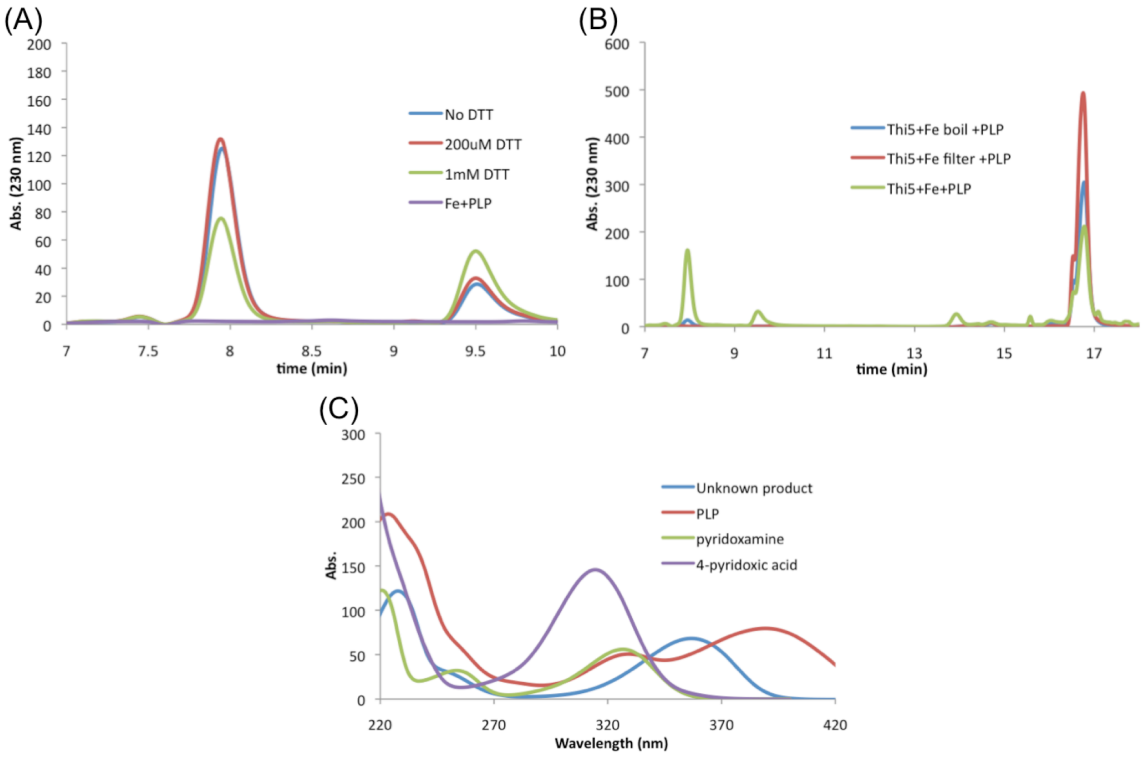


Figure 5-1. (A) The generation of a PLP-derived shunt product under low concentration of DTT or no DTT. (B) PLP-derived shunt product generated by active THI5p. (C) UV spectra of the PLP-derived shunt product and pyridoxal analogues.

The corresponding fraction was collected by HPLC and then analyzed by LC-MS. The $[M+H]^+$ of this compound was 245.0 Da (Figure 5-2A). Interestingly, the MS/MS spectrum suggested this compound had a phosphate group. This was confirmed

by the treatment of calf intestinal phosphatase (CIP). The retention time of this compound in HPLC analysis shifted after CIP treatment (Figure 5-2B). This suggested that this compound might have a PLP like structure. We used ^{15}N-THI5p to test whether there was a MW shift of this compound and found that the MW increased by 1 Da (Figure 5-2A), but the PLP added was not ^{15}N-labeled, implying that ^{15}N-THI5p donated one ^{15}N.

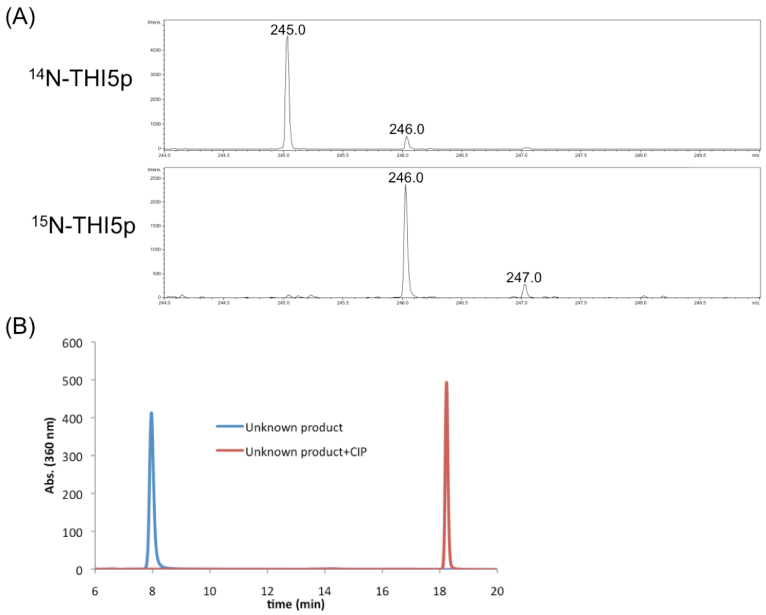


Figure 5-2. (A) PLP-derived shunt product generated by ^{14}N- and ^{15}N-THI5p. (B) PLP-derived shunt product has a phosphate group.

Therefore, the structure was proposed as PLP-nitrile **5-6** (Scheme 5-1). The reference compound without the phosphate group was synthesized by Dr. Dmytro

Fedoseyenko. The reference compound co-migrates with this compound after CIP treatment (Figure 5-3).

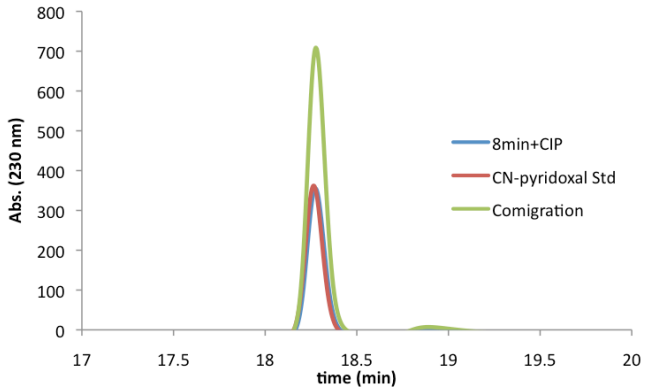
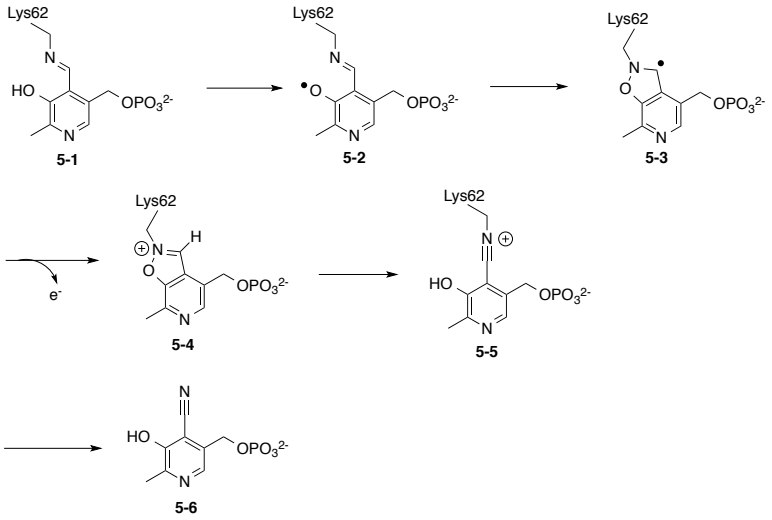


Figure 5-3. The co-migration experiments for the PLP-derived product and reference compound.

The formation of PLP-nitrile **5-6** could support that HMP-P synthesis is initiated by radical generation (Scheme 5-1). **5-1** undergoes $1e^-$ oxidation to generate **5-2**. Phenolic radical of **5-2** attacks the imine to form **5-3**. After $1e^-$ oxidation, **5-4** is formed to undergo deprotonation to generate **5-5**. A nucleophile attacks **5-5** to generate **5-6**.



Scheme 5-1. The proposed mechanism of PLP-nitrile **5-6** formation

C199 and M320 oxidation

In the THI5p crystal structure, the active-site structure shows that PLP forms an imine with Lys62, located near His66. In the sequence alignment of THI5p homologs, the sequence LGCCCFC (residues 193~199) is highly conserved. In the *in vitro* studies, any of these four cysteines after mutagenesis would abolish HMP-P production, indicating the essential nature of the existing interactions. After the THI5p full reaction, the inactive THI5p was treated with iodoacetamide to protect the cysteines followed by trypsin digestion. LC-MS analysis identified one corresponding peptide with +48 Da (Figure 5-4), suggesting cysteine oxidation. MS/MS analysis identified that the C199 thiol becomes a sulfonic acid. These findings were analyzed by MALDI-TOF MS and MS/MS analysis.

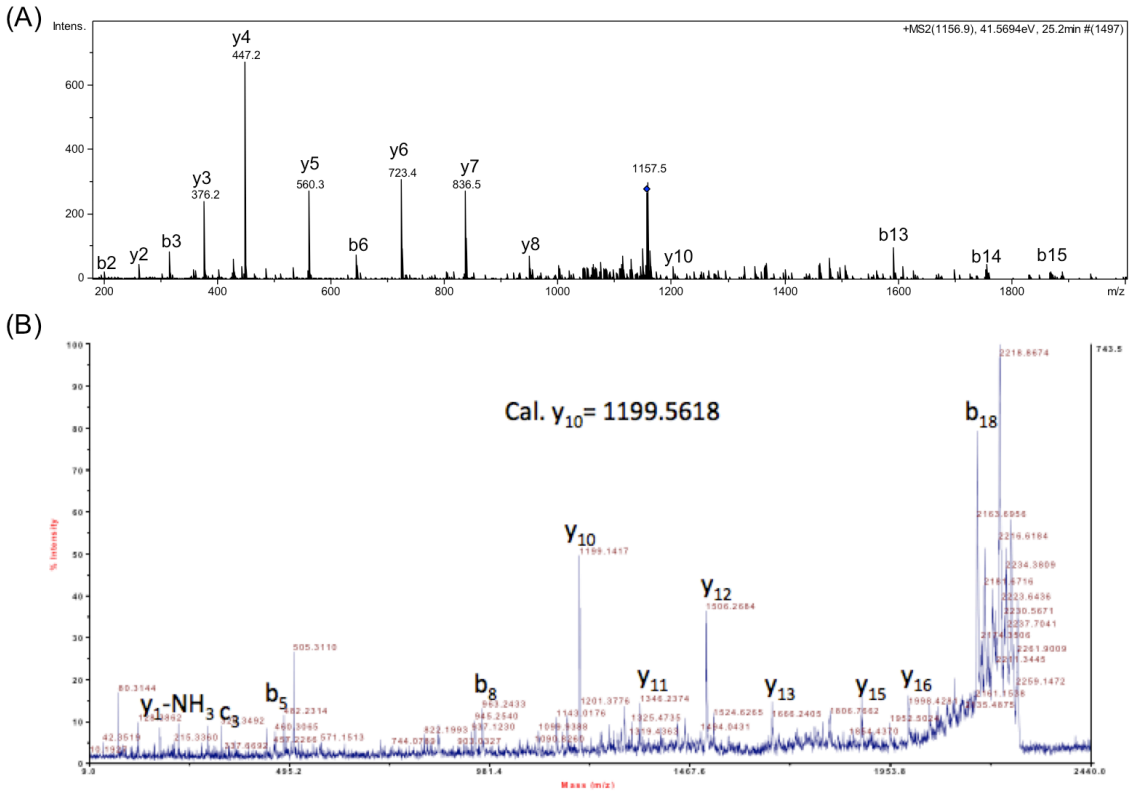


Figure 5-4. Identification of C199 oxidation: (A) ESI-MS/MS confirms the peptide sequence. (B) MALDI-TOF MS/MS identifies y_{10} ion corresponding to C199 oxidation.

A peptide containing M320 was also found showing a 16 Da (Figure 5-5) increase in MW. Under an atmosphere of $^{18}O_2$, this peptide showed a +2 Da shift derived from ^{18}O incorporation. Unfortunately the MS/MS spectrum could not identify the position, but M320 was the only residue that could be oxidized.

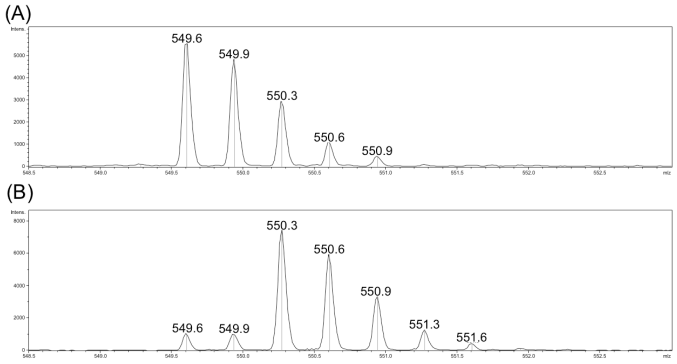


Figure 5-5. The MS analysis of peptide corresponding M320 oxidation under atmosphere of $^{16}O_2$ (A) and $^{18}O_2$ (B).

Intact protein molecular weight of the inactive THI5p

We have discussed three modifications in inactive THI5p. They are at His66, Cys199, and Met320 which were identified by the peptide analysis. If all the MW shifts are added, the MW of THI5p would increase by about 70 Da if HMP-P formation resulted in these modifications. During the course of THI5p studies, the assay conditions were improved to generate more HMP-P. Because the activity of THI5p is better, it resulted in fewer protein species after HMP-P formation. Finally, the intact protein MW of inactive THI5p was measured and showed an increase of 70 Da than the THI5p calculated MW (Figure 5-6).

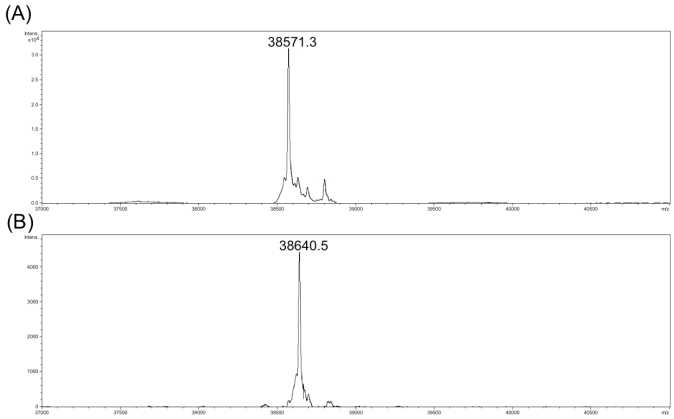


Figure 5-6. The intact MW analysis of apo-THI5p (A) and inactive THI5p after HMP-P synthesis (B).

Identification of 2-oxo-histidine

During the course of peptide analysis, one peptide derived from the control assay (THI5p+Fe no DTT) had an attractive MS/MS spectrum (Figure 5-7) with similar signals as the a_2, b_2 and y_2 ions of the modified peptide. Interestingly, the MW of this peptide is 16 Da more than the native peptide. The MS/MS analysis of this peptide indicated the sequence comprises residues 63-71, and that His66 has a modification increasing it by 16 Da. To test whether +16 Da derived from oxygen gas, the assay was incubated under an atmosphere of $^{18}O_2$, which showed ^{18}O incorporated into the modified His66.

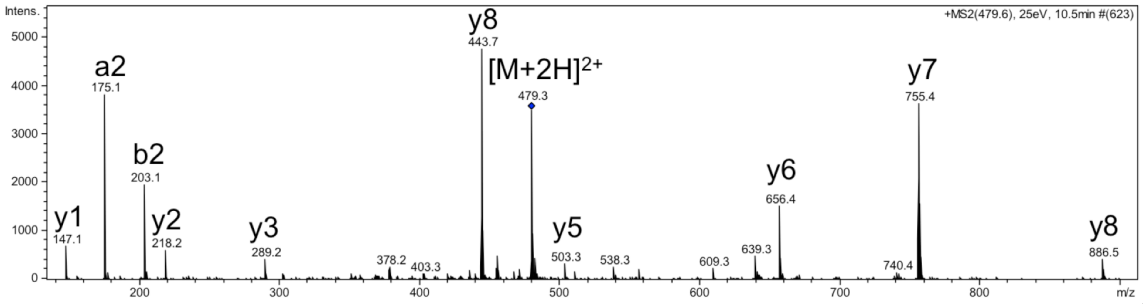


Figure 5-7. The MS/MS analysis of the peptide containing 2-oxo-histidine.

The oxidation of histidine on the protein can occur under oxidative stress conditions. One of the structures is 2-oxo-histidine derived from the hydroxyl radical [27,28]. In order to test the presence of 2-oxo-histidine, after the assay THI5p was hydrolyzed for 24 h at 105 °C in the presence of 6 M HCl solution with 40 mM DTT (Figure 5-8). After removing the solvent, the sample was redissolved in 0.1% acetic acid and subjected to LC-MS analysis for the co-migration. The reference compound was made by Dr. Dmytro Fedoseyenko. The EIC analysis showed good co-migration, and the MS/MS spectrum of this compound was identical to that reported in the literature [29].

Ortho-phthalaldehyde (OPA) was used to derivatize 2-oxo-histidine to conduct the co-migration. Since the mixture of hydrolysate contained a lot of small molecules, the fraction containing 2-oxo-histidine was collected by HPLC. After removing the solvent, OPA derivatization was performed on the purified hydrolysate and the reference compound. The mixture was analyzed by HPLC monitored at 340 nm (Figure 5-9). The HPLC analysis suggested that the purified hydrolysate contained 2-oxo-histidine.

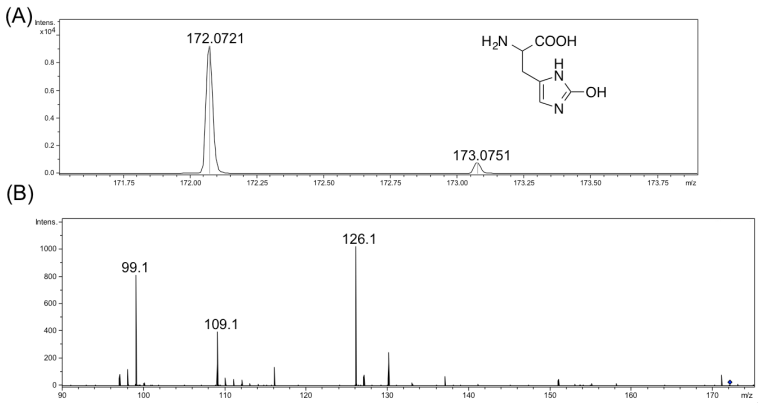


Figure 5-8. The MS (A) and MS/MS (B) analysis of 2-oxo-histidine in THI5p hydrolysate.

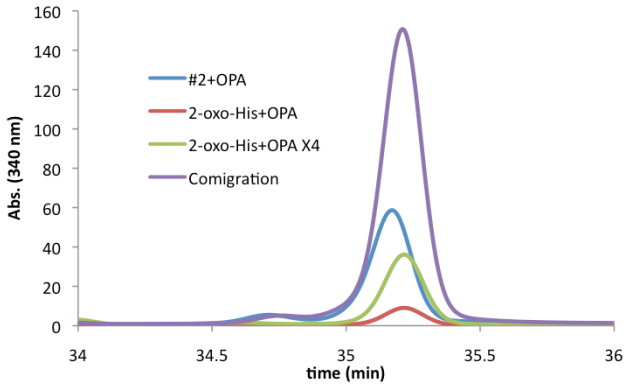


Figure 5-9. The co-migration experiments by *ortho*-phthalaldehyde (OPA) derivatization to identify 2-oxo-histidine generation in THI5p.

The identification of 2-oxo-histidine constitutes indirect evidence of the presence of H_2O_2 which catalyzes the decarboxylation of the C3-fragment to generate glyoxylate. H_2O_2 generation is proposed to occur when O_2 reacts with a di-ferrous.

Identification of C199 oxidation and 2-oxo-histidine in ScTHI5p

Like the His66-derived keto-acid, do C199 and His66 become oxidized *in vivo*? Since *Sc*THI5p was overexpressed in *S. cerevisiae*, *Sc*THI5p digested by trypsin was analyzed by LC-MS.

For C199, the corresponding peptide with an increase of 48 Da was identified by EIC. The MS/MS spectrum confirmed the sequence (Figure 5-10), but could not identify the oxidized residue. Thus, it was implied that cysteine oxidation occurs, but was not proven conclusively.

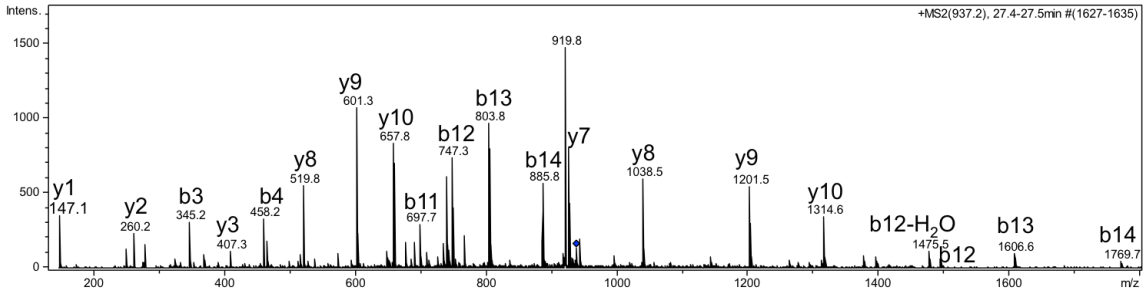


Figure 5-10. The MS/MS analysis of the peptide containing possible C199-sulfonic acid.

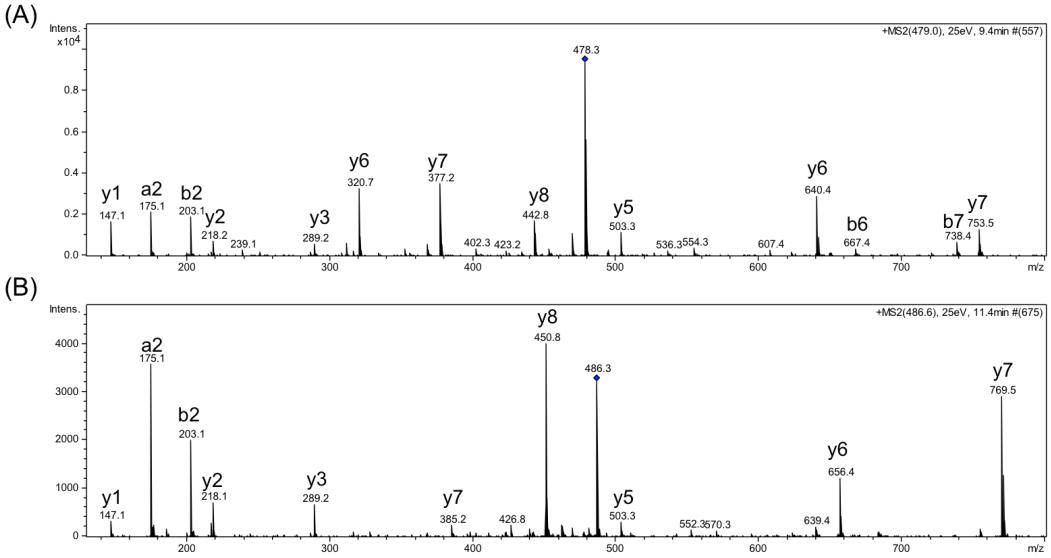


Figure 5-11. The MS/MS analysis of the *S. cerevisiae* peptide containing His66 (A) and His66-derived 2-oxo-histidine (B).

For His66, the corresponding peptide with +16 Da was identified by EIC. The MS/MS analysis confirmed the sequence and identified +16 Da at His66 (Figure 5-11). But this modification is not part of HMP-P formation. I have no explanation for the function of this modification. It is also possible that it is just an artifact.

H66C oxidation

His66 could receive the 2-oxo-histidine modification under non HMP-P production assays. If His66 was mutated to a residue that was easily oxidized, what modification would occur? To test this, the THI5p H66C was mutated to a cysteine using the same conditions as for the generation of 2-oxo-histidine. Cysteine can be oxidized to

sulfenic, sulfinic, or sulfonic acid. Sulfonic acid is the most stable product for cysteine oxidation. Therefore, the first candidate peptide was to look for a 48 Da increase using EIC. However, the corresponding peptide was not found. Interestingly, the peptide with +32 Da was identified suggesting the generation of sulfinic acid (Figure 5-12), which is unexpected given its lesser stability. At this moment, these data cannot be interpreted.

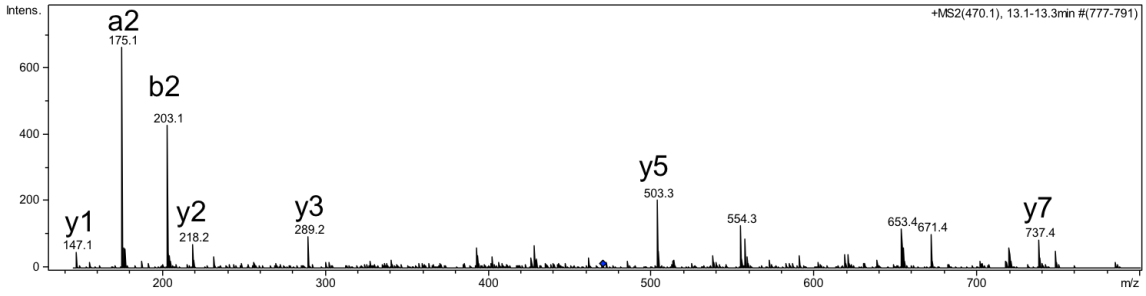


Figure 5-12. The MS/MS analysis of the peptide containing H66C with +32 Da.

Conclusion

In this chapter, I have discussed several modifications of THI5p. C199 and M320 oxidation occur as part of HMP-P formation. Moreover, C199 is conserved in all THI5p homolog and C199 mutants cannot generate HMP-P, suggesting that C199 plays an important role in HMP-P formation. It has been proposed that C199's function involves the electron transfer that initiates HMP-P formation, but a definitive answer awaits further investigation.

The formation of 2-oxo-histidine suggests the presence of H_2O_2 which could support decarboxylation of the C3-fragment. The most important discovery is the

identification of PLP-nitrile. The unexpected nitrogen incorporation is derived from THI5p from the generation of a phenolic radical involving HMP-P formation.

CHAPTER VI

IDENTIFICATION OF BOUND INTERMEDIATES ON THI5 PROTEIN

Introduction

The mechanistic investigation of HMP-P synthesis catalyzed by THI5p is complicated due to His66's involvement in the reaction. Though HPLC analysis does not identify any shunt products in the small molecule pools, it does not rule out formation of shunt intermediates on the protein. Molecular weight analysis of the intact protein does not provide a definitive identification because there are other modifications and PLP bound to the protein. Moreover, very small modifications to the entire protein could not be identified due to sensitivity.

In this chapter, the radioactive approach was developed to identify shunt intermediates on THI5p. After trypsin digestion of protein, Fe-NTA chromatography was used to enrich phosphopeptides containing shunt intermediates which were analyzed by LC-MS and MS/MS. Preliminary characterization of the peptides were completed to make proposed structures of shunt intermediates.

Experimental methods

Preparation of ^{32}P-pyridoxal-5'-phosphate (^{32}P-PLP)

12 μL of 10 mM pyridoxal was added to 50 μL of PdxK (pyridoxal kinase), 24 μL of 10 mM ATP and 5 μL of γ-^{32}P-ATP (10 Ci/mmol). Then 2 μL of 0.1 M $MgCl_2$

was added to initiate the phosphorylation. The reaction mixture was incubated at 37 °C for 2 h. Finally, the mixture was filtered with a 10 kDa cut-off filter. The filtrate was the stock solution of ^{32}P-PLP for further uses.

Detection of bound intermediates in the THI5p full reaction assay by phosphoimager

100 µL of 600 µM THI5p or variants were anaerobically pre-incubated with 2 molar equivalent of $Fe(NH_4)_2(SO_4)_2$ in an ice bath for 30 min. The mixture was anaerobically desalted using a Bio-Spin 6 column (Bio-Rad) pre-equilibrated with 100 mM Tris-HCl, 5 mM DTT, pH 7.5. Subsequently, 1 µL of ^{32}P-PLP stock solution and 2 molar equivalents of PLP were added. The mixture was incubated at room temperature for 3 h. After the reaction finished, 10 µL of the reaction mixture was added into 10 µL of SDS buffer (20% glycerol, 4% SDS, 100 mM Tris, pH 7.4). Then the mixture was incubated at 90 °C for 5 min. After the mixture was cooled, 10 µL of 500 mM methoxyamine, pH 7.0 was added and incubated at room temperature for 30 min. Finally, 150 µL of 100 mM Tris, pH 7.4 was added to dilute the mixture. 10 µL of the diluted mixture was used to run SDS-PAGE. The gel was dried by the gel dryer system (Bio-rad) and placed in the exposure cassette with a phosphoscreen for 1 h. The phosphoscreen was analyzed using a Typhoon scanner (GE health life science).

Small scale phosphopeptide enrichment

After the THI5p full reaction assay was completed, the mixture was buffer-exchanged to 20 mM ammonium bicarbonate by Bio-Spin 6 column (Bio-Rad). Trypsin

was added to digest THI5p at 37 °C overnight at a 1:50 ratio of trypsin to THI5p. Then the solution was dried by a speedvac concentrator (Thermo Scientific). The Fe-NTA phosphopeptide enrichment kit (Pierce) was used to enrich the resulting peptide mixture, followed by LC-MS analysis.

Large scale phosphopeptide enrichment

After THI5p purification, 3 mL of 1 mM THI5p was incubated with 1 mL of 10 mM PLP. Then the mixture was desalted under anaerobic conditions using a 10DG column (BioRad) pre-equilibrated with 100 mM Tris-HCl buffer, 30% glycerol, pH 7.5. Three molar equivalents of $Fe(NH_4)_2(SO_4)_2$ were added into the resulting THI5p solution in an ice bath for 30 min. The mixture was anaerobically desalted using a 10DG column (Bio-Rad) pre-equilibrated with 100 mM Tris-HCl, 5 mM DTT, pH 7.5. Then the mixture was separated into 500 µL aliquots. Finally, 50 µL of air saturated buffer (100 mM Tris-HCl, pH 7.5) was added aerobically and incubated at room temperature for 3 h.

Following incubation, 3 mL of the reaction mixture was desalted by a 10DG column (Bio-Rad) pre-equilibrated with 20 mM ammonium bicarbonate, pH 8.0. Trypsin was added into the resulting THI5p solution and incubated at 37 °C overnight (the ratio of trypsin to THI5p was 1:50). Then the mixture was incubated at 60 °C for 30 min followed by cooling to room temperature. Trypsin was again added into the THI5p solution and incubated at 37 °C for another 8 h. Finally, the solution was dried by lyophilization.

The IMAC column for phosphopeptide enrichment was prepared using NTA-Agarose (Qiagen). First, 20 mL of 0.2 M $FeCl_3$ flushed NTA-Agarose followed by 20 mL of water to remove unbound Fe^{3+}. Finally, the Fe-IMAC column was equilibrated with the loading solution (6% acetic acid, pH 3.5). After lyophilization the sample was dissolved in the loading solution and vortexed vigorously followed by centrifugation at 5000 rpm for 10 min. The supernatant was injected into Fe-IMAC column and washed with the washing buffer (90% loading buffer and 10% acetonitrile), followed by equilibration with the loading buffer. Finally, the bound peptides were eluted by 200 mM $NH_4H_2PO_4$, pH 4.4. The eluted peptides were dried by lyophilization and dissolved in 0.1% formic acid for LC-MS analysis or HPLC collection.

Results and discussion

Detection of bound intermediates by phosphoimager

In the proposed mechanism, most of the proposed intermediates are on the protein instead of existing as separate small molecules because His66 is the substrate for HMP-P formation. Therefore, the ^{32}P-radiological experiments were applied to identify any shunt intermediates on the protein using phosphoimaging. ^{32}P-PLP was made from pyridoxal with γ-^{32}P-ATP catalyzed by pyridoxal kinase (PdxK). ^{32}P-PLP was added into the THI5p or variants assays. After the reaction, the protein was heat-denatured in SDS buffer followed by the addition of methoxyamine. Without the addition of methoxyamine, ambiguous signals were derived from ^{32}P-PLP forming an imine with

Lys62. Finally, the signals analyzed by phosphoimager were compared between the full and control reactions.

The imaging results showed that the signal in THI5p-WT full reaction assay was stronger than the other three control reactions (Figure 6-1A), which suggested the presence of shunt intermediates. Moreover, the full reactions in other THI5p variants also had stronger signals (Figure 6-1B).

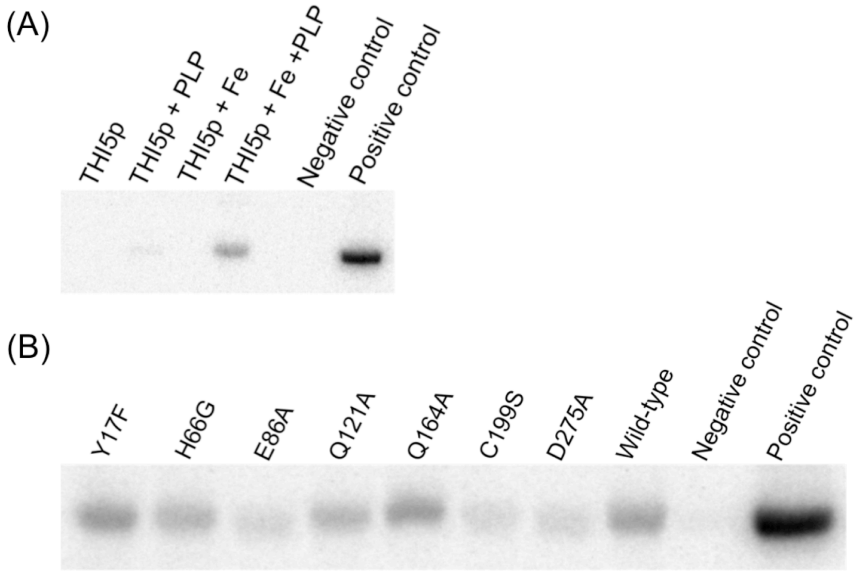


Figure 6-1. The phosphoimaging analysis of the THI5p-WT assays (A) and THI5p variant full reactions (B). Negative control: THI5+PLP followed by methoxyamine and $NaBH_4$ reduction. Positive control: THI5p+PLP followed by $NaBH_4$.

Identification of shunt intermediates on the THI5p wild-type

The phosphoimaging results suggested the presence of shunt intermediates in the THI5p-WT full reaction. However, we did not have structural information and did not know the identity of the modified residue. We decided to identify the residues first. Because the phosphoimager signals were derived by ^{32}P, we knew the shunt intermediates had a phosphate group, so phosphopeptide enrichment was used [30]. This is widely used method in proteomic research for phosphorylated proteins. Because phosphopeptide signals are not as strong as normal peptides, phosphopeptides must be enriched for MS identification. For THI5p, the enrichment by Fe-IMAC worked better than TiO_2. It was used to enrich the phosphopeptide from the digestion sample of THI5p-WT from the full reaction [31]. The control reaction was THI5p+PLP followed by the same enrichment procedure. Both samples were analyzed by LC-MS. After comparing the LC-MS analysis from these two samples, five peptides were identified with shunt intermediates (Figure 6-2 to 6-6). The MS/MS analysis of all five peptides showed two peptide sequences. The sequence of two peptides (A and B) were from Ala63 to Lys71 and contained His66. The sequence of three peptides (C, D, and E) were from Ala57 to Lys71 and contained both Lys62 and His66. These results were promising because they supported the proposed mechanism as any intermediate would have a covalent bond with His66. Moreover, four out of these five peptides had a phosphate group based on the loss of 98 Da fragment in the MS/MS analysis.

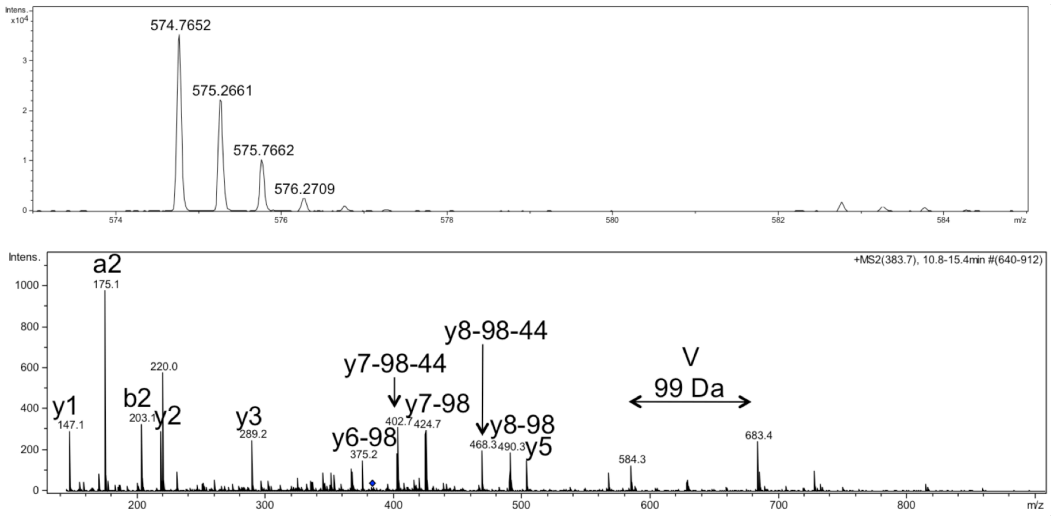


Figure 6-2. The MS and MS/MS analysis of peptide A.

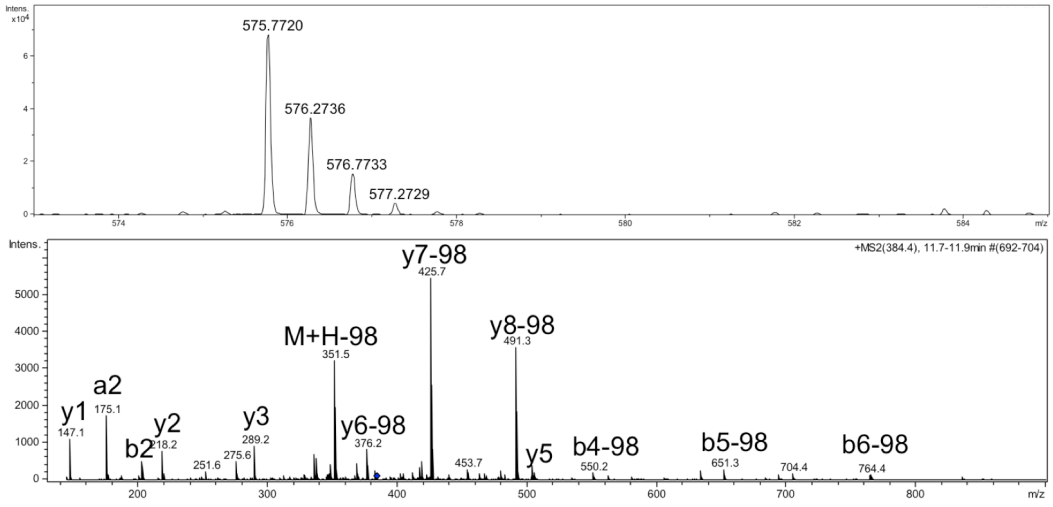


Figure 6-3. The MS and MS/MS analysis of peptide B.

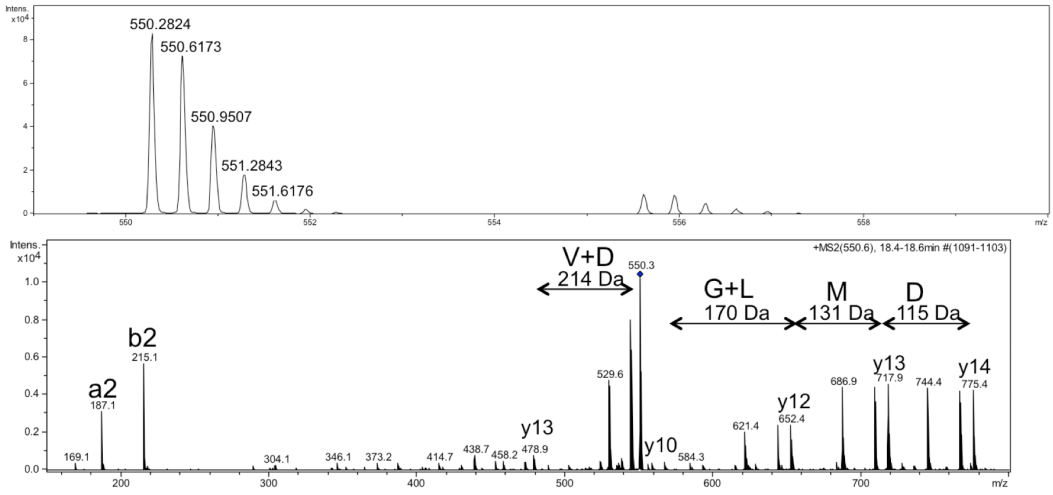


Figure 6-4. The MS and MS/MS analysis of peptide C.

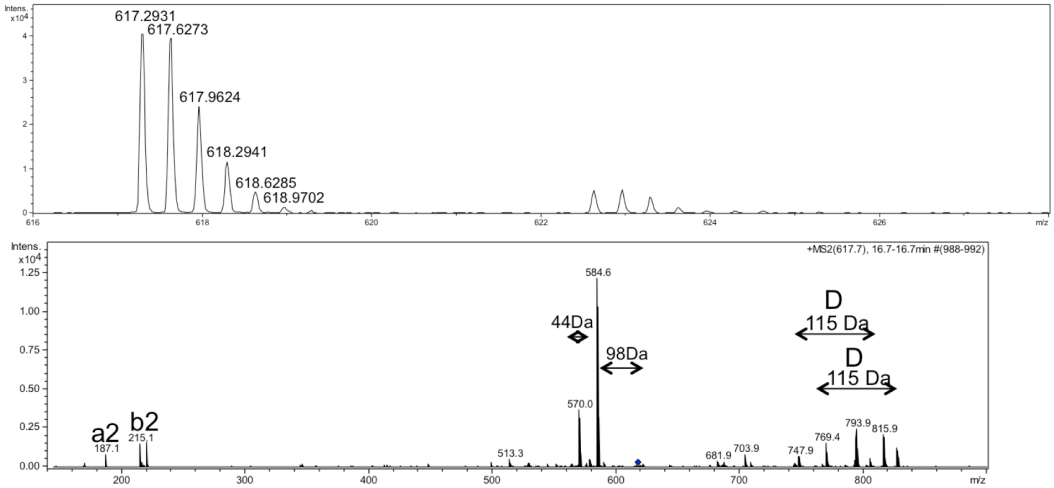


Figure 6-5. The MS and MS/MS analysis of peptide D.

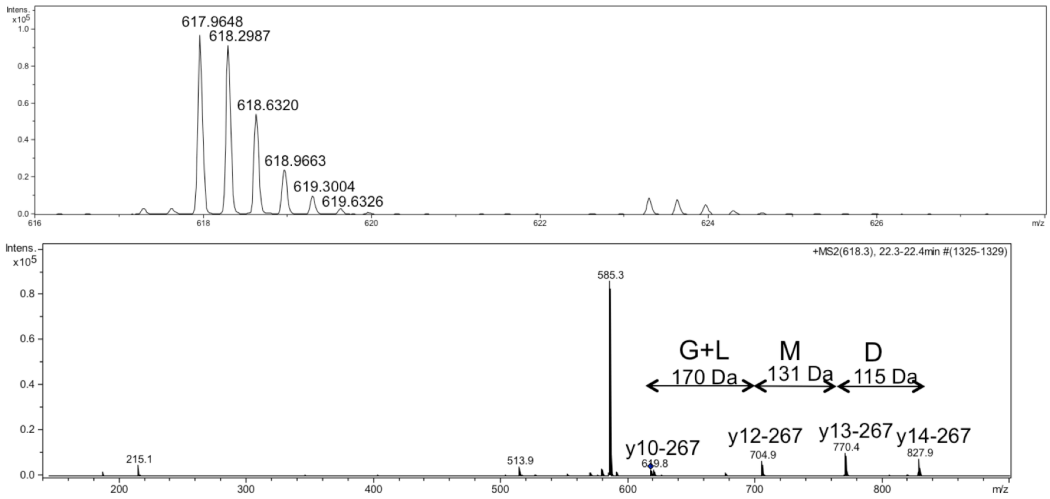


Figure 6-6. The MS and MS/MS analysis of peptide E.

Identification of the five peptides derived from the enzymatic reaction

The peptide analysis suggested these five peptides might be candidates to study. But in order to confirm these peptides derived from THI5p reaction, 4',5'-^{13}C-PLP was used to set up the assays. Our hypothesis is that the peptides would have ^{13}C incorporation if they were derived from shunt reactions. The MS analysis of these five peptides confirmed that ^{13}C incorporated into all of peptides (Figure 6-7 to 6-11). Peptides A, B, and C had one ^{13}C incorporation. Peptides D and E had two ^{13}C incorporations. The results further confirmed that these five peptides contain shunt intermediates derived from the THI5 enzymatic reaction.

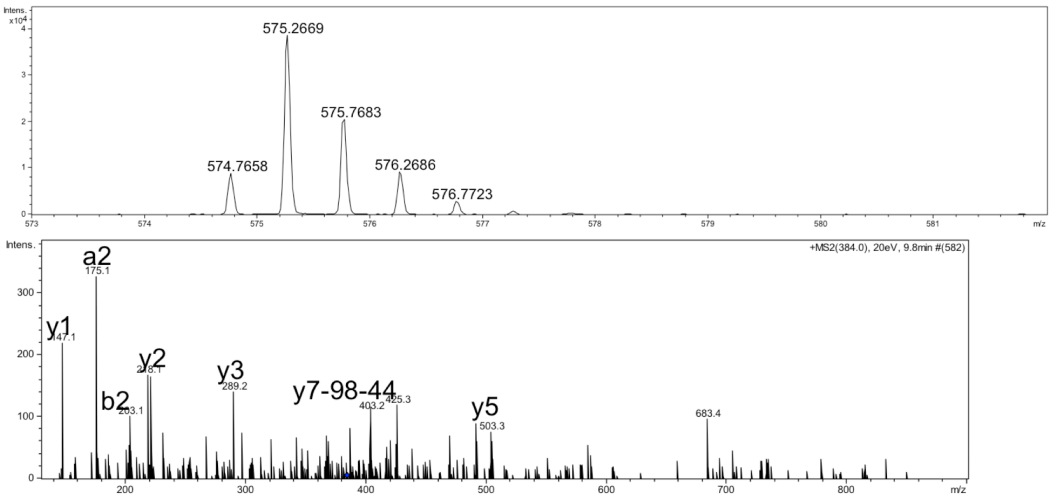


Figure 6-7. The MS and MS/MS analysis of peptide A with one ^{13}C incorporation.

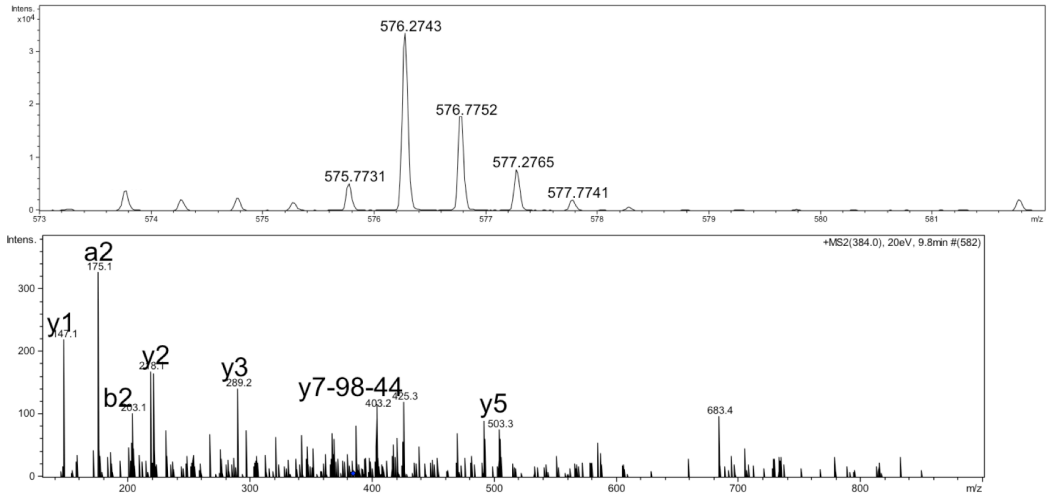


Figure 6-8. The MS and MS/MS analysis of peptide B with one ^{13}C incorporation.

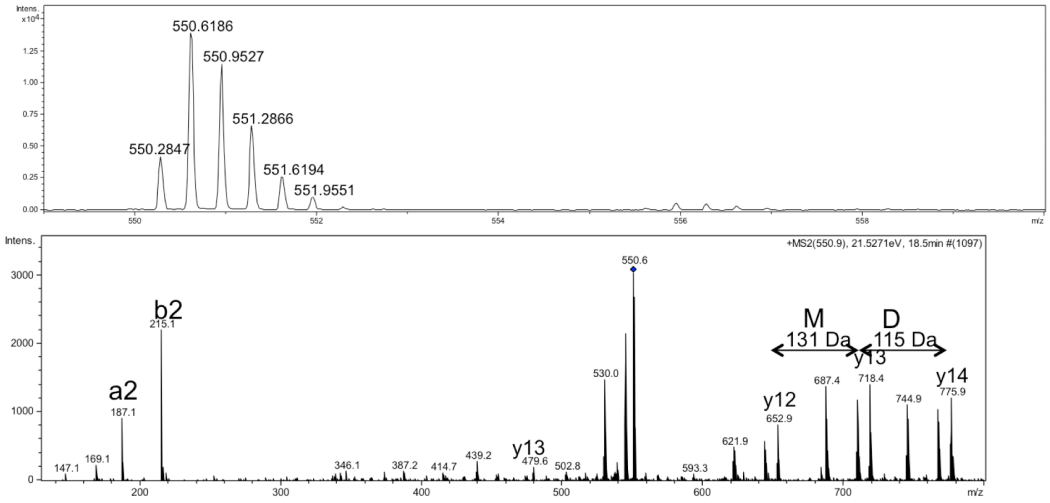


Figure 6-9. The MS and MS/MS analysis of peptide C with one ^{13}C incorporation.

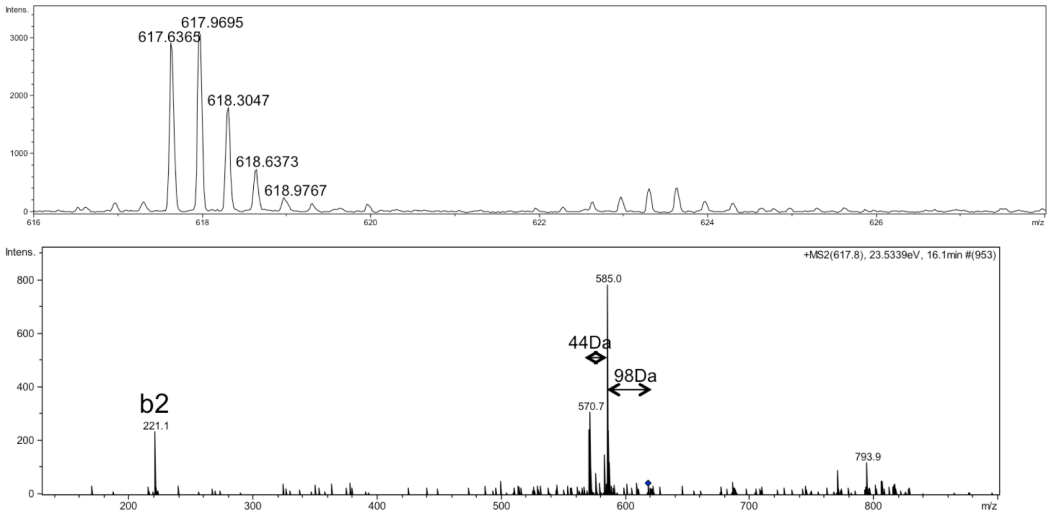


Figure 6-10. The MS and MS/MS analysis of peptide D with two ^{13}C incorporations.

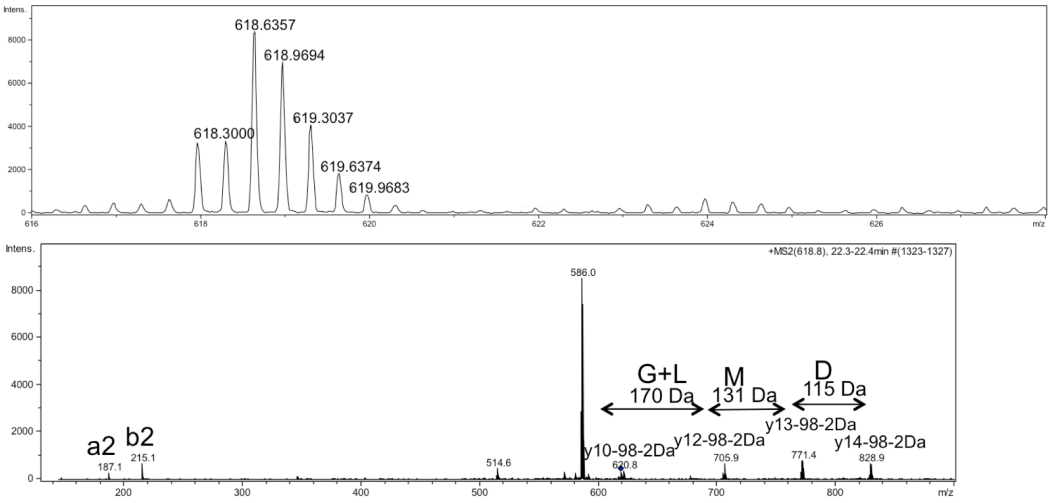


Figure 6-11. The MS and MS/MS analysis of peptide E with two ^{13}C incorporations.

Identification of the number of nitrogen at His66

In order to solve the structure of the shunt intermediates, I wanted to investigate whether the imidazole of His66 was intact or if N=C-N was donated to generate HMP-P. To discover this the number of nitrogen atoms can be identified by ^{15}N-THI5p. ^{15}N-THI5p was used to set up the assays. After peptide analysis (Figure 6-12 to 6-16), ^{15}N-peptides A and B had both an increase of 12 Da compared with ^{14}N-peptides A and B. Based on the peptide sequence, there are 12 nitrogen atoms, thus two nitrogen atoms of the His66 imidazole were retained in the shunt intermediates.

Moreover, ^{15}N-peptides D and E had a 19 Da shift compared with ^{14}N-peptides D and E. Based on the peptide sequence, there are 19 nitrogen atoms, and two nitrogen atoms of imidazole at His66 and the nitrogen of Lys62 were retained in the shunt intermediates.

However, ^{15}N-peptide C had only a 17 Da shift compared with ^{14}N-peptide C. The peptide sequence indicates 19 nitrogen atoms meaning that two nitrogen atoms were lost during the reaction. I propose that two nitrogens from the imidazole of His66 were donated and the imidazole ring did not intact in the shunt intermediate.

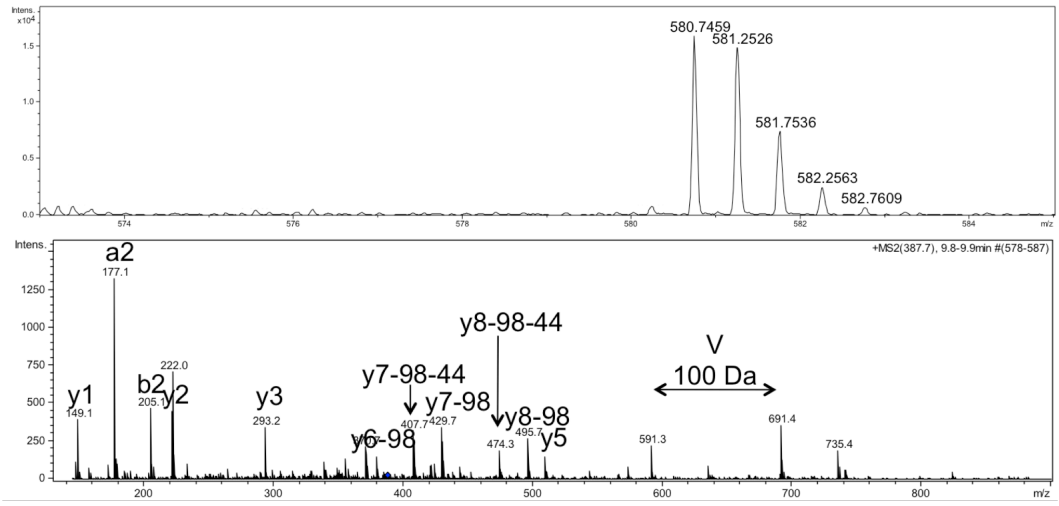


Figure 6-12. The MS and MS/MS analysis of ^{15}N-peptide A.

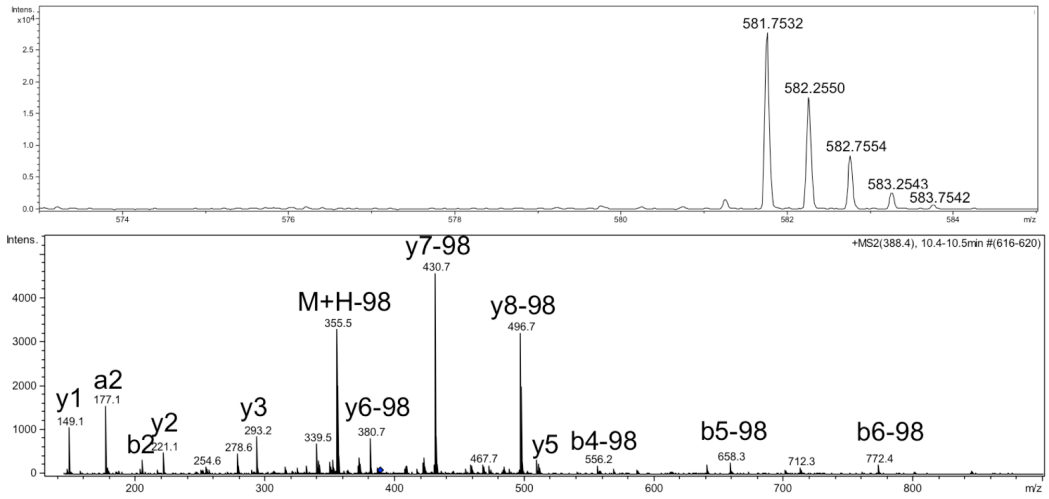


Figure 6-13. The MS and MS/MS analysis of ^{15}N-peptide B.

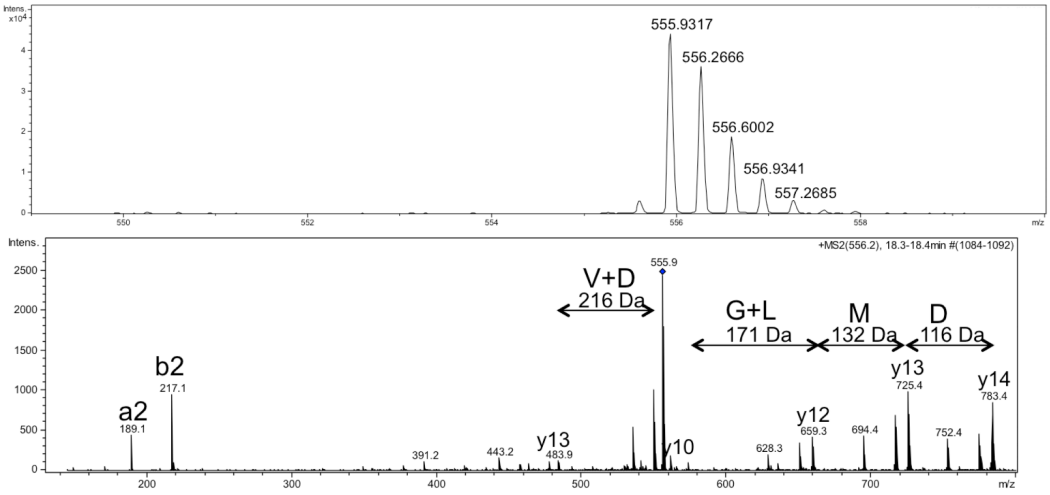


Figure 6-14. The MS and MS/MS analysis of ^{15}N-peptide C.

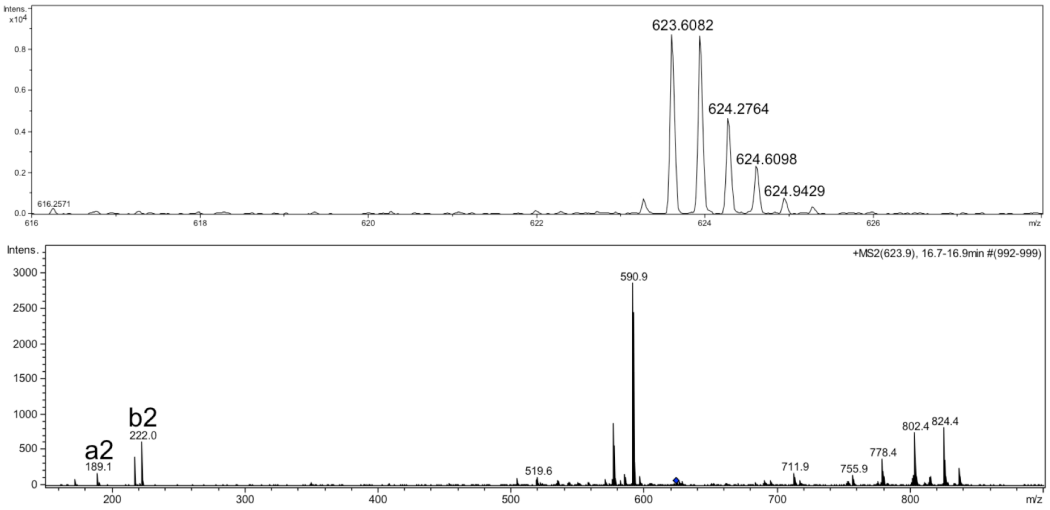


Figure 6-15. The MS and MS/MS analysis of ^{15}N-peptide D.

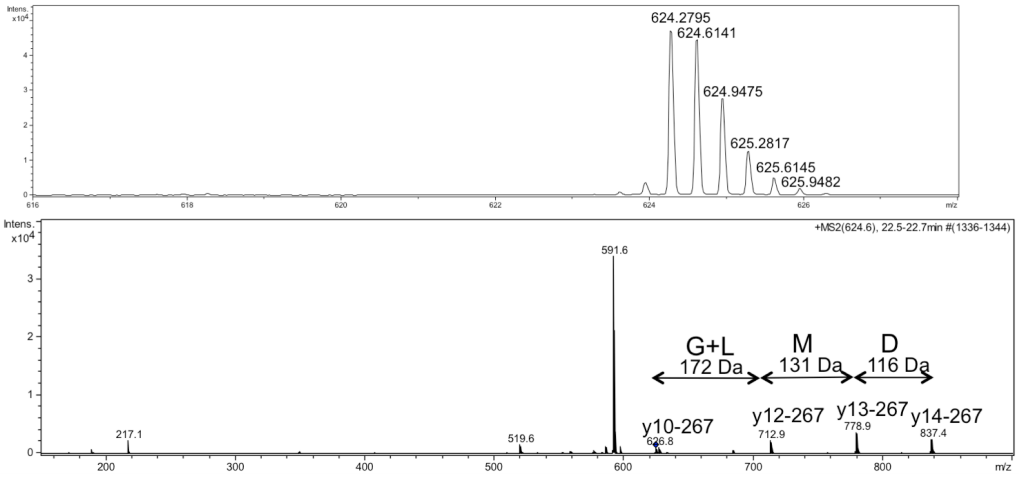


Figure 6-16. The MS and MS/MS analysis of ^{15}N-peptide E.

Identification of the carboxylic acid on shunt intermediates

During investigation of the structures of shunt intermediates, the presence of carboxylic acid was proposed. Therefore, a mixture of peptides A and B was isolated by HPLC. After drying the sample, anhydrous 3 M HCl in MeOH was added and stirred at room temperature for 2 h to undergo esterification. The mixture was dried and analyzed by LC-MS (Figure 6-17 and 6-18). Based on MS analysis, both peptides' MW increased by 28 Da corresponding to two methyl ester formations. Except for the C-terminal carboxylic acid, there are no carboxyl amino acid in peptides A and B and they must have come from the shunt intermediates.

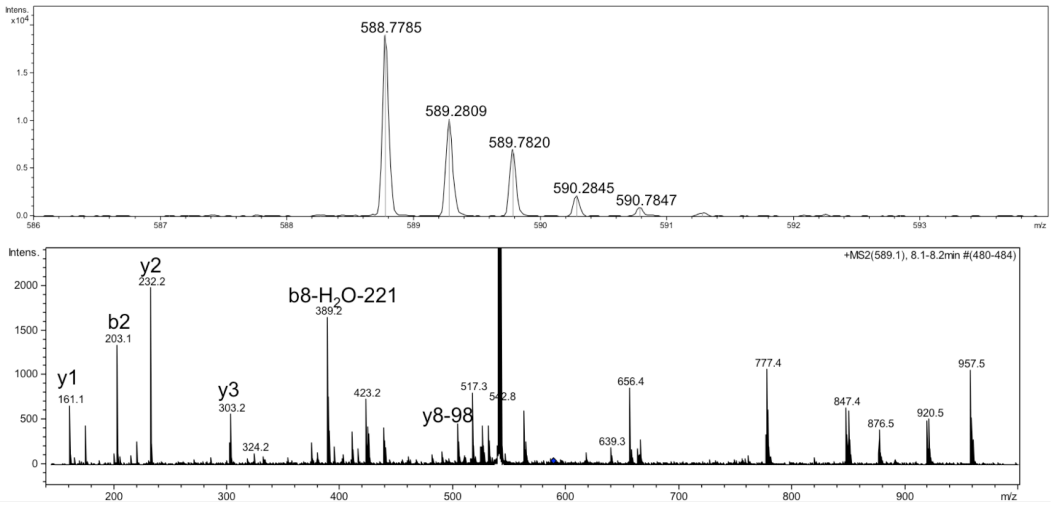


Figure 6-17. The MS and MS/MS analysis of peptide A after esterification.

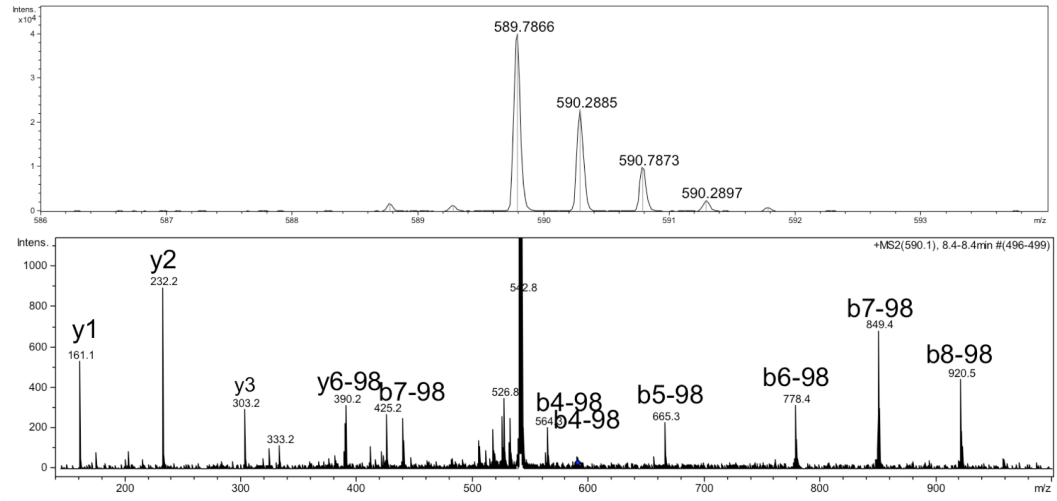


Figure 6-18. The MS and MS/MS analysis of peptide B after esterification.

The identification of the carboxylic acid raised a question about oxygen incorporation. The assay was set up under an atmosphere of $^{18}O_2$. The MS analysis of peptides A and B indicated two oxygen incorporations (Figure 6-19 and 6-20). Furthermore, peptides C and D showed three oxygen incorporations (Figure 6-21 and 6-22). The results were very similar to the oxygen incorporation of keto-acid. Although the experiment using a mixture of $^{16}O_2$ and $^{18}O_2$ was not conducted, the results were predicted to be that the two oxygen atoms derive from two different oxygen gas.

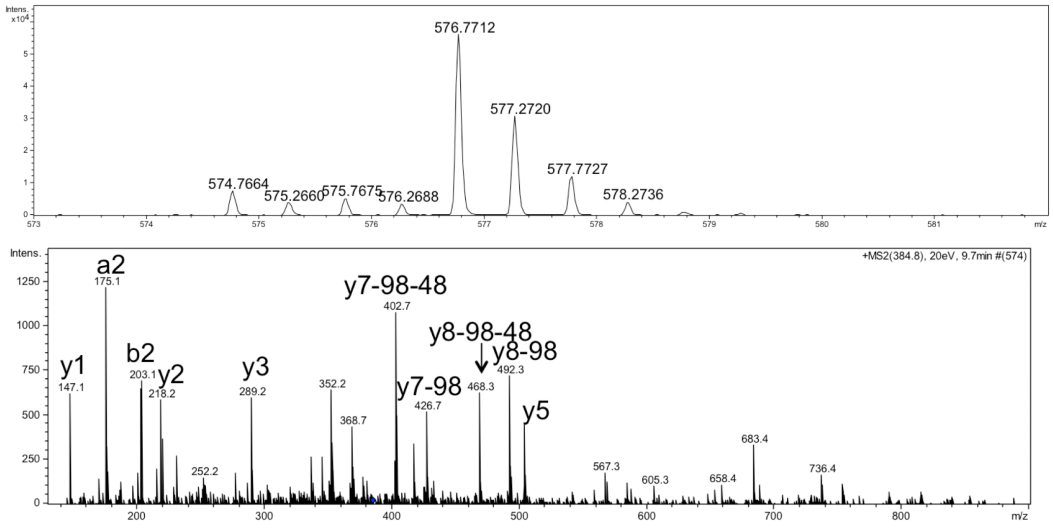


Figure 6-19. The MS and MS/MS analysis of peptide A with two ^{18}O incorporations.

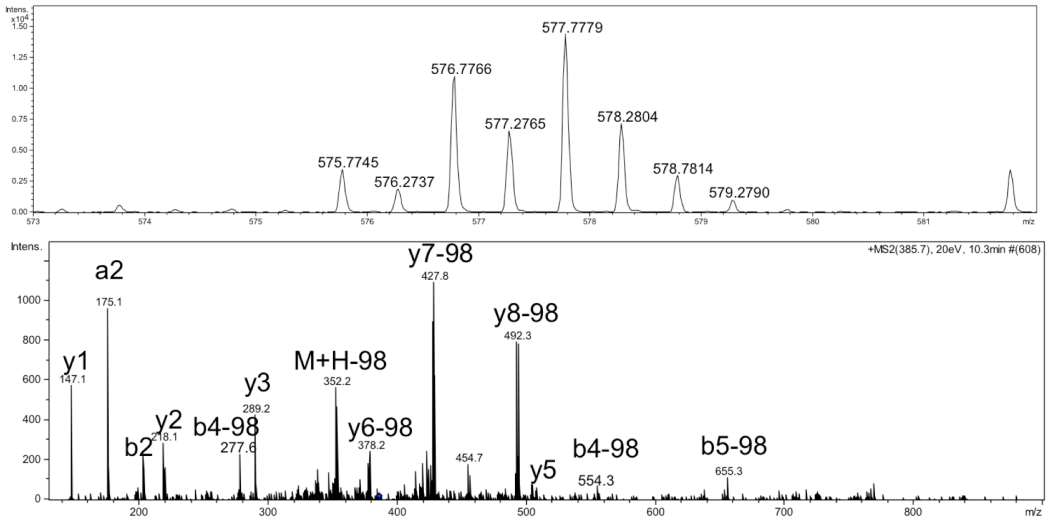


Figure 6-20. The MS and MS/MS analysis of peptide B with two ^{18}O incorporations.

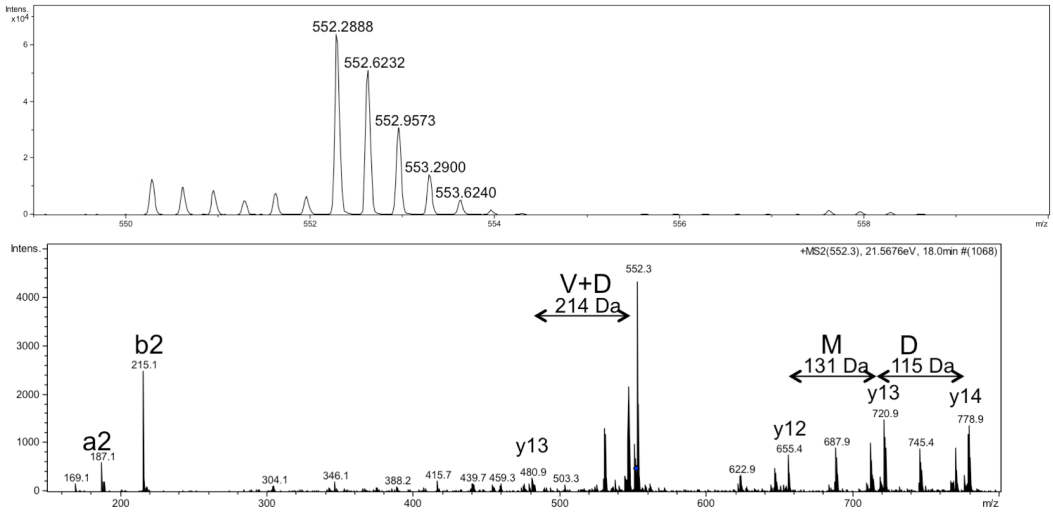


Figure 6-21. The MS and MS/MS analysis of peptide C with three ^{18}O incorporations.

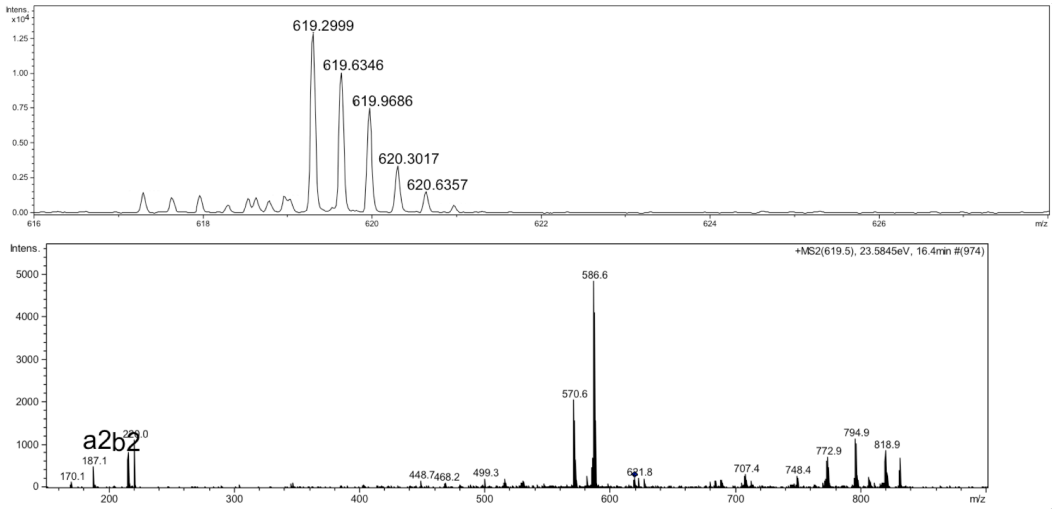


Figure 6-22. The MS and MS/MS analysis of peptide D with three ^{18}O incorporations.

Identification of +58 Da on Lys62

There are two peptide sequences with shunt intermediates. Peptides A and B comprise Ala63 to Lys71. Peptide A is +207 Da and peptide B is +209 Da compared to the original peptide. Based on the MS/MS spectra, both have a phosphate group. Interestingly, peptides D and E also have a phosphate group identified by the MS/MS spectra. They comprise residues Val57 to Lys71 and likely contain the same shunt intermediates at His66 that were in peptides A and B, but also include a modification at Lys62. If the modifications at His66 are identical, then peptides D and E both had an increase of 58 Da after subtracting the MW of the original peptides with 207 or 209 Da.

To identify the 58 Da structure at Lys62, the THI5p H66G mutant was used to investigate. After the reaction, the protein was digested with trypsin and analyzed by LC-MS (Figure 6-23). The EIC analysis was used to find a peptide between Val57 and Lys71 showing +58 Da. Based the MS and MS/MS analysis, the peptide can be identified. Interestingly, this structure was Lys62 with glyoxylate after an imine reduction. Although the mechanism of this modification is unknown, structural identification could be achieved. Glyoxylate was added to THI5p H66G followed by $NaBH_4$ reduction. The resulting protein was digested followed by co-migration experiments. The retention time and MS/MS spectra were identical, confirming the structure.

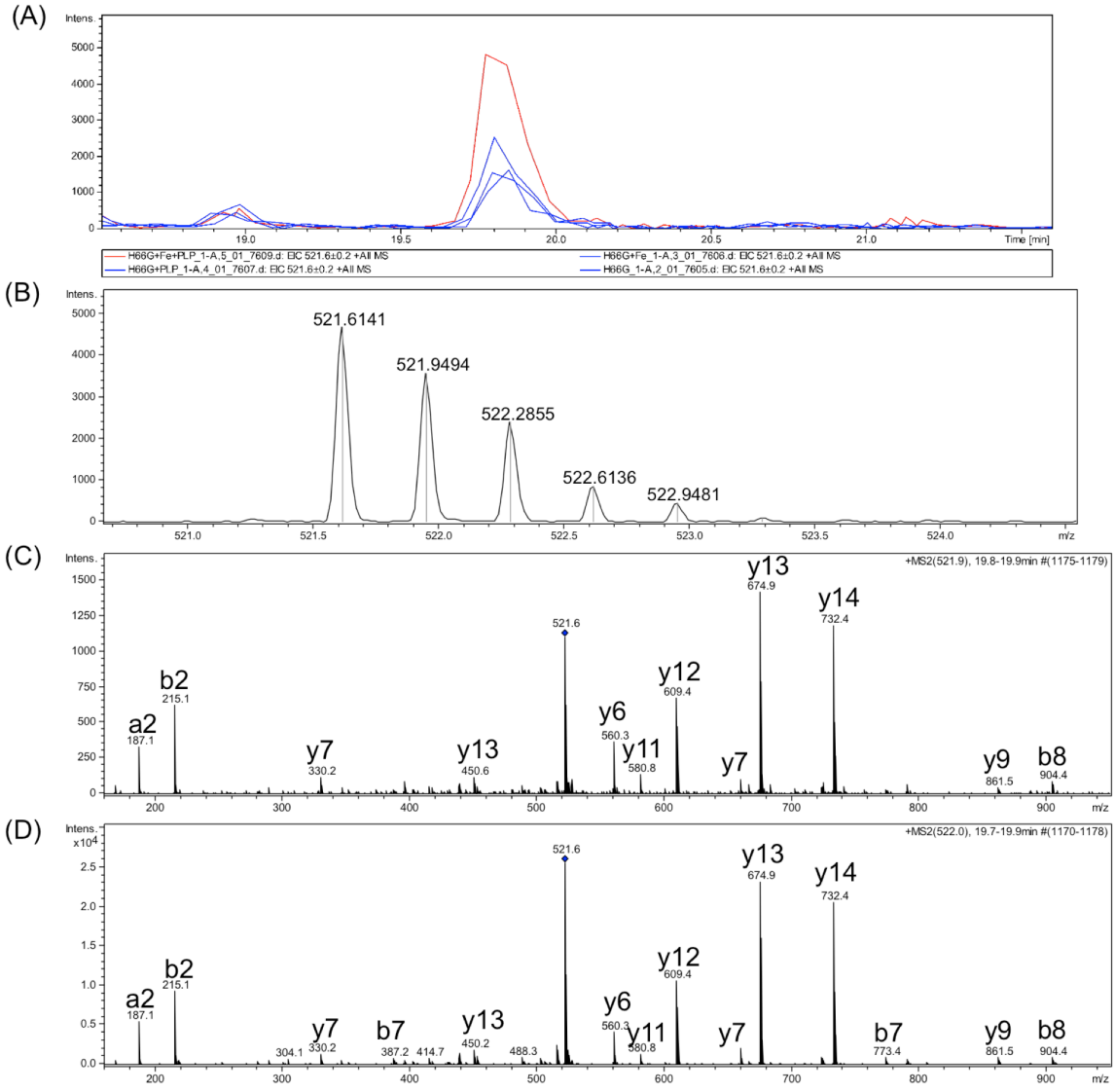


Figure 6-23. (A) The EIC analysis of the peptide with +58 Da. (B) The MS analysis of the peptide with +58 Da. (C) The MS/MS analysis of peptide with +58 Da. (D) The MS/MS analysis of reference peptide generated by THI5p H66G+glyoxylate with $NaBH_4$ reduction.

I also attempted to characterize peptide C. The sequence of peptide C is the same as peptide D and E and I assumed that peptide C had +58 Da at Lys62. After subtracting 58 Da and the original peptide MW, the MW shift was 5.96 Da. This MW shift is the result of the histidine-derived keto-acid. To characterize this peptide, the digested peptides were treated with $NaBH_4$. The EIC analysis identified a peptide with +2 Da and the MS/MS spectrum confirmed the correct peptide sequence (Figure 6-24). This characterization indirectly solved the structure of peptide C.

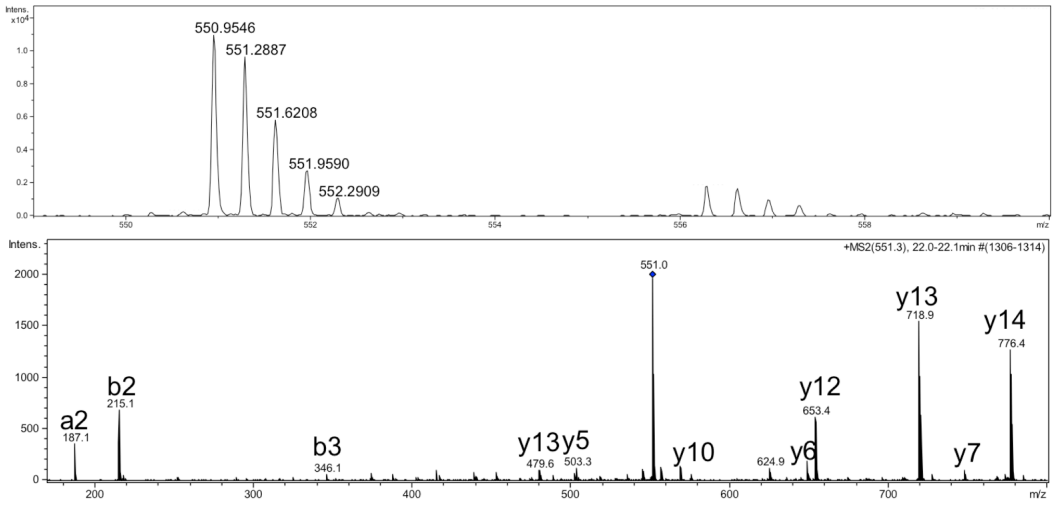


Figure 6-24. The MS and MS/MS analysis of peptide C after $NaBH_4$ reduction.

Conclusion

Based on the available information, the proposed structures of all five peptides are shown (Figure 6-25). Further research is required to characterize the structures.

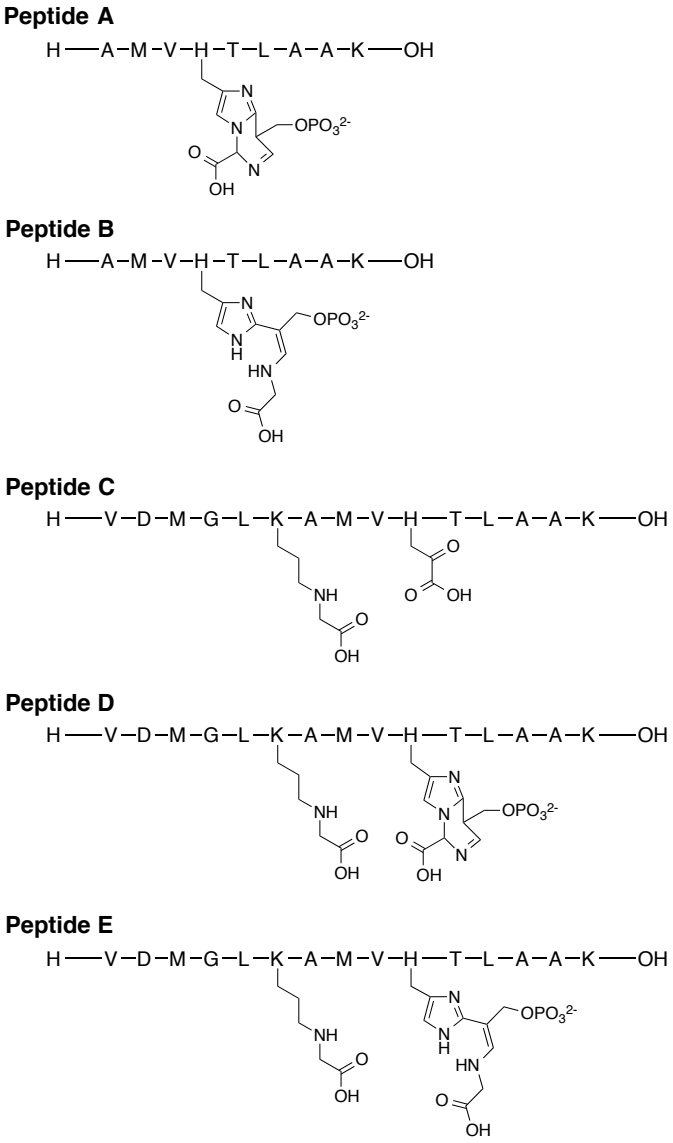


Figure 6-25. The proposed structure of five peptides.

CHAPTER VII

PRELIMINARY BIOPHYSICAL INVESTIGATION OF THI5 PROTEIN

Introduction

The crystal structure of THI5p H66G with iron was obtained by Dr. Michael K. Fenwick in Dr. Ealick's lab (Cornell University). The preliminary result of this structure showed that THI5p is a di-iron protein. Notably, there were no histidines among the residues coordinating to the metal. Typically non-heme iron dependent enzymes have at least one histidine as the metal ligand [32,33], such as in the well-known di-iron enzymes, methane monooxygenase [34] and ribonucleotide reductase R2 [35], which each have one histidine for each iron. The absence of histidine presents the possibility that THI5p has a new iron coordinating environment to conduct HMP-P formation.

In this chapter, the oxygen consumption for HMP-P formation was measured and the preliminary EPR investigation of THI5p was conducted.

Experimental methods

Anaerobic overexpression of THI5p

E. coli BL21(DE3) cells containing the *THI5* gene in pET28b was grown in fermentor with nitrogen gas purge (the flow rate is 1 L/min). 10 L of growth medium contained 250 g of LB, 9.3 L of water, 500 mL of buffer stock (1 M Na_2HPO_4, 1 M KH_2PO_4, and 0.5 M $(NH_4)_2SO_4$), 250 mL of sugar stock (20% glucose and 20%

glycerol), 20 mL of 1 M $MgSO_4$, and 250 mL of 1 M sodium fumarate. Every stock solution was autoclaved separately. The medium contained kanamycin (100 μg/mL) with 50 rpm of stir at 37 ^oC until the OD_{600} reached 0.6. At this point, protein overexpression was induced with IPTG (final concentration of 1 mM), and cell growth was continued at 22 °C for 16 h. The cells were harvested and the protein was purified anaerobically.

Determination of iron amount

The active THI5p was prepared with the standard procedure in the glove box. 10 μL of THI5p solution was added into 490 μL of water followed by the addition of 100 μL of 2 M HCl and 100 μL of saturated ammonium acetate. The solution was incubated at 100 °C for 3 min. The resulting solution was centrifuged at 16000 rpm for 5 min. 600 μL of the supernatant was removed and mixed with 20 μL of 10 mM ferrozine and 20 μL of 100 mM ascorbic acid. After 10 min incubation, the absorbance was measured at 562 nm (the extinction coefficient of iron ferrozine complex at 562 nm is 27.9 $mM^{-1}cm^{-1}$).

Oxygen consumption measurement

The Oxygraph system (Hansatech instruments) was used to measure the oxygen concentration of the reaction mixture. The instrument installation and the calibration followed the manual description. The sample chamber contained 500 μL of air saturated buffer. 100 μL of the THI5p solution prepared in the glove box was injected into the

sample chamber. The oxygen concentration was recorded by the instrument. Because the THI5p solution was anaerobic, the final oxygen concentration was multiplied by 5/6 for the initial reading. Similarly the final concentration of THI5p would be 1/6 of the original. The calculation used the final concentration after mixing the anaerobic THI5p solution and air saturated buffer.

TCEP oxidation in the THI5p C199A assays analyzed by ^{31}P-NMR

200 µL of THI5p C199A solution was anaerobically added into 500 µL of buffer (100 mM Tris-HCl, 2 mM TCEP, pH 7.5). The mixture was stirred at room temperature under $^{18}O_2$ or $^{16}O_2$ for 2 h. The mixture was put into an Amicon 10 KDa cut-off filter and centrifuged at 14000 rpm. 500 µL of filtrate was mixed with 60 µL of D_2O and 3 µL of 6 M HCl. ^{31}P-signal of the solution was measured by Bruker 400 MHz.

Results and discussion

Oxygen consumption measurement of THI5p wild-type

In the previous chapters, the oxygen incorporation of the His66-derived keto-acid and PLP by-products were investigated using $^{18}O_2$. Based on these results, the proposed mechanism of HMP-P formation included several oxygen requirements. Therefore, the change of oxygen concentration was measured by the Oxygraph system.

The active THI5p concentration was presumably determined by half the iron concentration because of the di-iron site. The iron concentration was measured by

Ferrozine assay. The change of oxygen concentration was divided by the active THI5p concentration to get the ratio corresponding to the total oxygen consumption per THI5p.

The full reaction assay of THI5p-WT was measured (Figure 7-1). The initial O_2 concentration was 250 µM. Because 100 µL of the THI5p oxygen free solution was added into 500 µL of air saturated buffer, the actual initial O_2 concentration was 208.3 µM. The final O_2 concentration was 70 µM. The consumed oxygen was 133.3 µM. Moreover, the concentration of iron on THI5p was 27 µM determined by a Ferrozine assay. Based on the di-iron enzyme model, the amount of active THI5p was assumed to be 13.5 µM. Finally, the calculated ratio of the consumed oxygen to the active THI5p was 10.2.

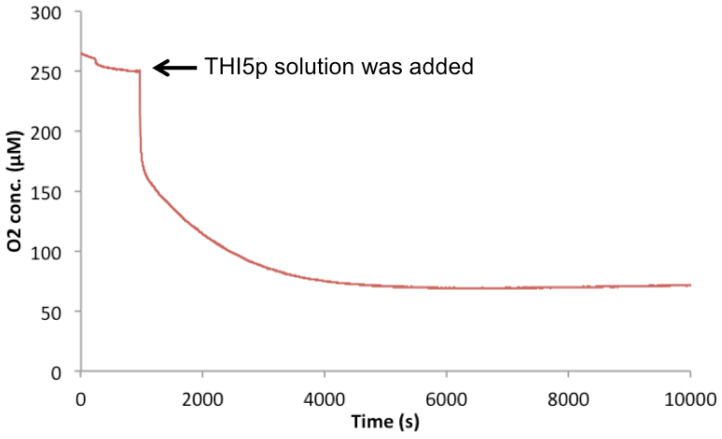


Figure 7-1. Measurement of oxygen consumption in THI5p-WT reaction.

This result suggested that the THI5p full reaction uses multiple O_2, which is different from any non-heme iron dependent oxygenase. Moreover, the proposed

mechanism showed four O_2 for HMP-P formation, and C199 and M320 oxidation might require another four oxygen gas molecules. However, the metal-site chemistry is unknown and might react with oxygen to conduct the reaction as in ribonucleotide reductase R2. The unconventional chemistry is an interesting feature of the metal-site.

Oxygen consumption measurement of THI5p C199A

Although the function of C199 is unclear, it is essential for HMP-P formation. Therefore, the measurement of oxygen consumption was conducted for the full reaction assay using the THI5p C199A mutant (Figure 7-2). Unexpectedly, this mutant consumed more oxygen than the wild-type to the point where it depleted all the oxygen gas in the reaction buffer. To test its reactivity, oxygen saturated buffer was added and monitored continuously. The result showed it consumed all the oxygen gas present in the buffer. This is currently perplexing, but it may be that C199 plays an important role in metal-site chemistry. The cysteine is also required for function as the mutant cannot generate HMP-P and the di-iron site likely reacts with oxygen to conduct the reaction.

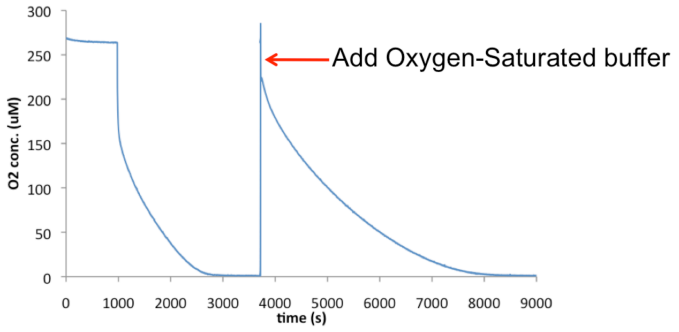


Figure 7-2. The measurement of oxygen consumption in the THI5p C199A reaction.

TCEP oxidation investigation using THI5p C199A

THI5p C199A reacted with oxygen gas catalytically, but was unable to generate the product, HMP-P. Presumably, the mutation prevented oxygen from incorporating on the protein. The intact MW of THI5p-C199A after the reaction also showed no obvious modification. The high oxygen gas consumption of this protein indicates the generation of a product, but what product is it?

If O_2 reacts with di-ferrous, it could generate hydrogen peroxide with the formation of di-ferric. Testing this hypothesis with the dye PO1 did not detect the formation of hydrogen peroxide [36].

Next, I tested for the reaction of Tris-(2-carboxyethyl) phosphine (TCEP) which can be oxidized. After the reaction, the protein was removed and the filtrate was analyzed by ^{31}P-NMR [37] (Figure 7-3). The solution was adjusted to be acidic to make the TCEP signal visible. In the ^{31}P-NMR spectra, the full reaction showed the formation of Tris-(2-carboxyethyl) phosphine oxide (TCEPO) compared with other two control reactions.

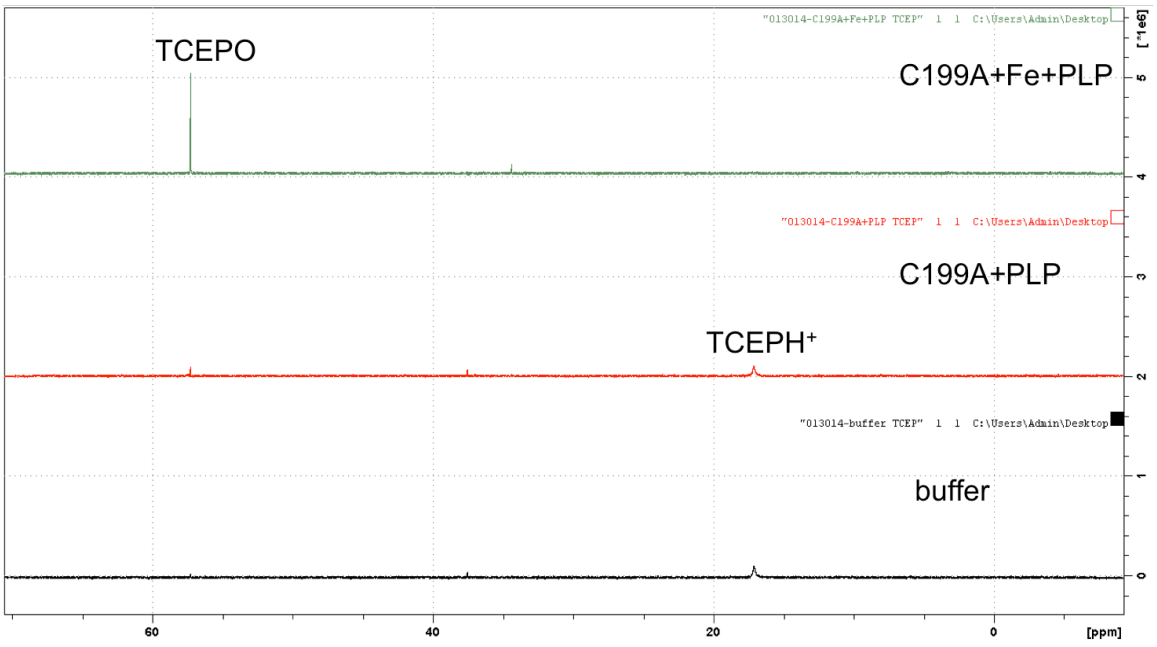


Figure 7-3. ^{31}P-NMR monitoring oxidation of TCEP in THI5p C199A assays.

However, there are two cysteine residues, C196 and C197, near the di-iron site that could form a disulfide bond. TCEP reacts with disulfide to regenerate free cysteine with formation of TCEPO. To differentiate between these two reactions, the assay was set up under an $^{18}O_2$ atmosphere. If TCEPO is ^{18}O-labeled, it suggests that the by-product of O_2 oxidizes TCEP. ^{31}P-NMR spectra showed TCEPO with ^{18}O incorporation (Figure 7-4). However, the ratio of ^{18}O- and ^{16}O-labeled TCEP was 0.12 (the expected ratio is 0.5).

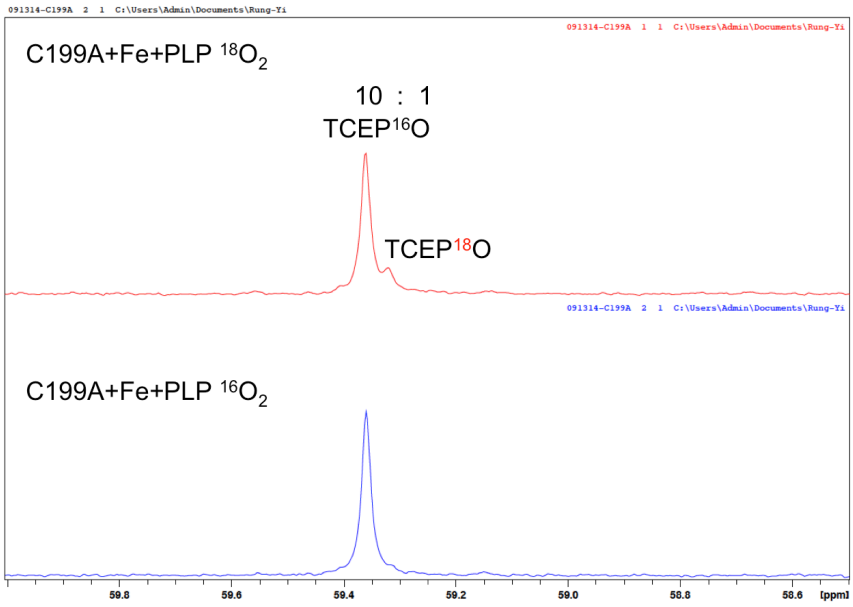


Figure 7-4. ^{31}P-NMR monitoring oxidation of TCEP in the THI5p C199A assays under atmosphere of $^{16}O_2$ (bottom) and $^{18}O_2$ (top).

The results did not match the proposed mechanism, but they could give some hints about the di-iron chemistry of THI5p. The oxidant derived from oxygen gas and di-iron might allow the formation of a disulfide bond between C196 and C197. TCEP regenerates free cysteines with formation of TCEPO. ^{18}O-TCEPO might derive from the shunt reaction where the oxidant reacts with TCEP before C196 and C197. Clearly the di-iron chemistry will require more advanced biophysical investigations before yielding conclusive results, such as EPR and Mössbauer spectroscopy.

UV and EPR investigation of THI5p+Fe with NaCN

In the di-iron enzyme after air exposure, there is a charge transfer band at 340 nm corresponding to di-ferric oxo-bridged [38]. The addition of sodium azide would result in the formation of a chromophore with broad absorption bands at 345 and 450 nm, as observed for azide complexes and numerous other proteins containing oxo-bridged di-iron clusters.

Since the preliminary crystal structure of THI5p with iron suggested that it is a non-heme di-iron enzyme, similar investigations were applied. Because the PLP UV-Vis spectrum overlaps below 400 nm, it would be difficult to interpret the data, hence the addition of sodium azide was tested to generate a UV-Vis shift. However, no new charge transfer band was generated around 450 nm after adding sodium azide at 1 M final concentration. The lack of change might be due to a different coordinating environment, so a stronger ligand, cyanide, was tested. After exposing the air, sodium cyanide was added to the THI5p+Fe solution and, surprisingly, the solution became pink within seconds. The solution remained pink despite desalting to remove any small molecules implying that cyanide coordinates to iron. The UV-Vis spectrum of this solution was measured and λ_{max} was about 720 nm (Figure 7-5). The known cyanide bridged iron complex is Prussian Blue with broad absorbance about 600 nm. The oxidation state of di-iron is Fe(II) and Fe(III). Furthermore, the solution color did not change as drastically if THI5p+Fe was not exposed to air before the addition of sodium cyanide. Finally, this pink protein sample was measured by EPR. The g value was 2.11. But the typical di-iron enzyme has the g value below 2.

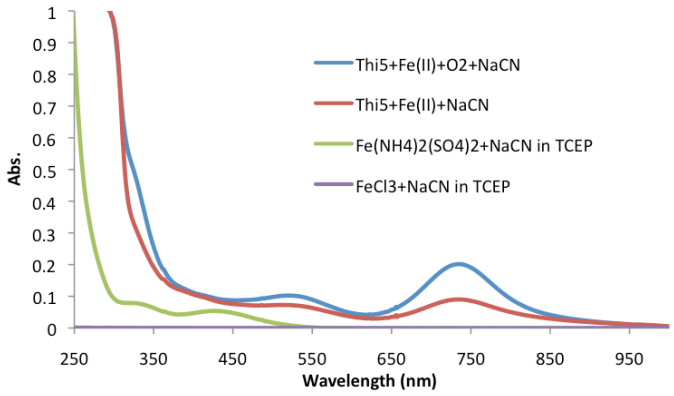


Figure 7-5. UV-Vis spectra of NaCN adding samples.

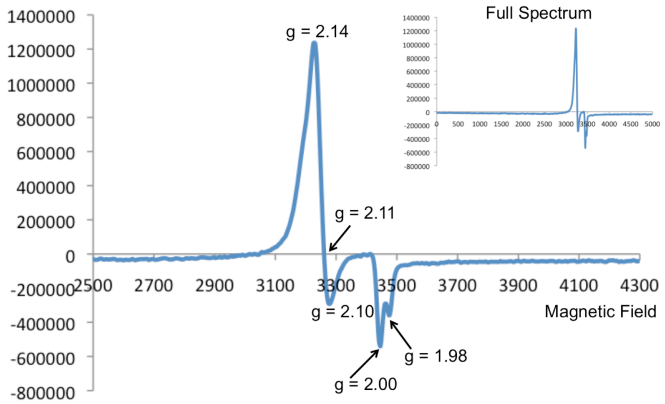


Figure 7-6. EPR measurement of THI5p+Fe(II)+O_2 with NaCN addition.

Conclusion

In this chapter, the preliminary investigation for the metal site was conducted. Although it could not fully interpret the bioinorganic chemistry, it was a starting point for the future investigation.

CHAPTER VIII

CONCLUSION AND PERSPECTIVE

Conclusion

Thiamin is a fascinating cofactor to investigate. In my first project with ThiG, the work was to use the reverse reaction to study later steps of bacterial thiazole biosynthesis. In my second project, the THI5p activity was reconstituted successfully. The chemical equation of HMP-P formation catalyzed by THI5p was completed (Figure 8-1).

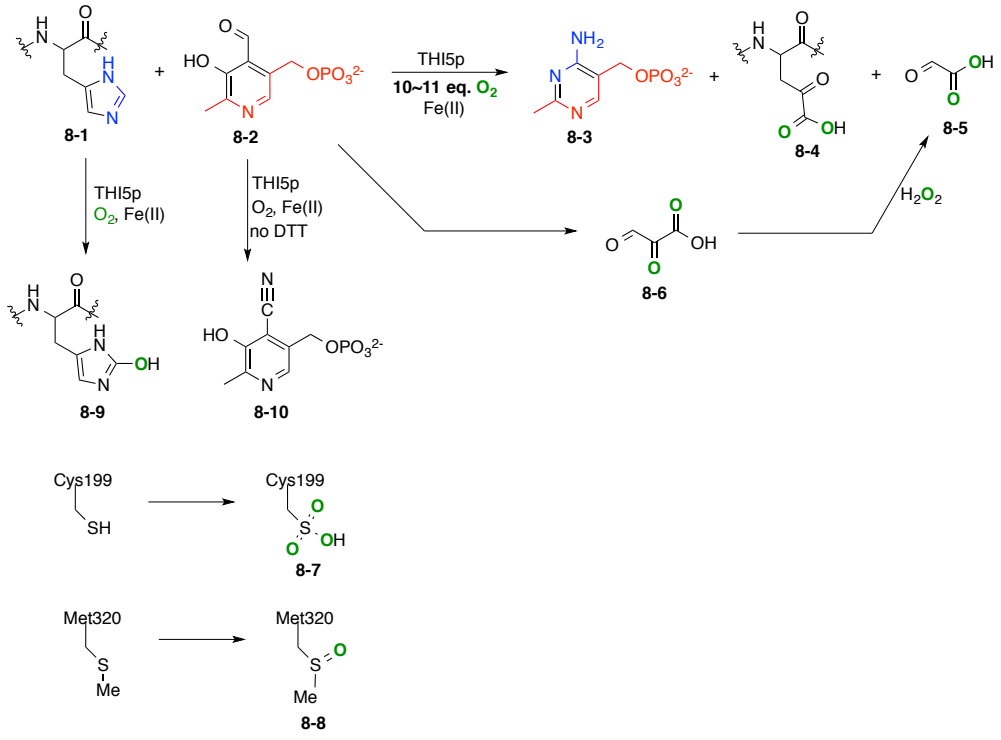


Figure 8-1. Summary of THI5p investigation

This enzymatic reaction requires 10 to 11 eq. of O_2 to generate HMP-P and oxygen incorporates into the His66-derived keto-acid **8-4** and glyoxylate **8-5**. Moreover, Cys199 and Met320 are oxidized after HMP-P formation. These are all unprecedented results. 2-oxo-histidine **8-9** and PLP-nitrile **8-10** were characterized to demonstrate the proposed mechanism.

Perspective

The mechanistic investigation of THI5p has been a challenge. In chapter 6, four peptides containing intermediates were identified. But LC-MS analysis provides only the peptide sequence information and the MW of the intermediates, it does not solve the structures. To characterize the intermediate structures, the purification of these peptides will be required followed by NMR analysis.

Next, the crystal structures of active THI5p+Fe and inactive THI5p are necessary to support investigation into the reaction mechanism. For inactive THI5p, it could compare with all of modification in the protein, e.g. the His66-derived keto-acid, Cys199 sulfonic acid, and Met320 sulfoxide (Figure 8-1). I have conducted preliminary crystallizing screening and formed some preliminary crystals, however, the best resolution obtained was about 3.5 Å. The active THI5p+Fe(II) crystal structure will be useful for biophysical investigations, e.g. EPR, because the simulation of the spectra requires structural information to yield an accurate interpretation. The tentative ligands for iron do not have any histidine which is present in all other non-heme iron dependent enzymes.

For the mechanistic investigation, PLP analogues could be applied to divert the reaction by the formation of shunt intermediates. Analogues **8-11** to **8-15** (Figure 8-2) made by Dr. Dmytro Fedoseyenko were tested. The preliminary results suggested that **8-11** and **8-15** might generate some shunt intermediates based on intact protein MW measurements. Based on these results, it will be interesting to use halogenated PLP at the C2 position, e.g. **8-16** and **8-17**, to generate more shunt intermediates.

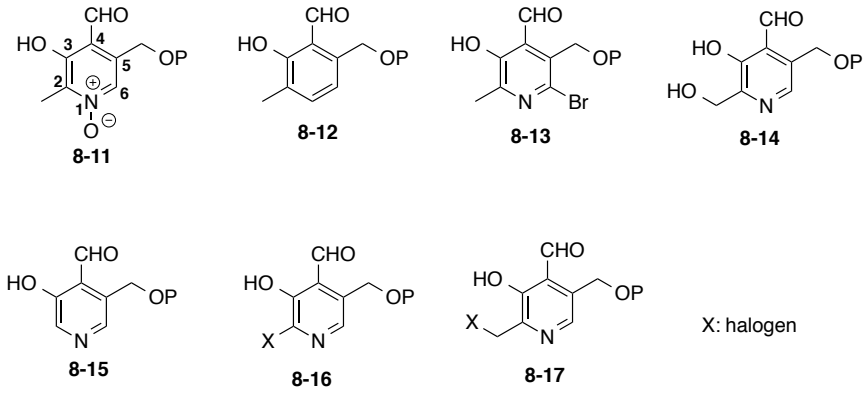


Figure 8-2. The PLP analogues for THI5p mechanistic investigations. The substrates were synthesized by Dr. Dmytro Fedoseyenko.

Other experiments were also conducted to try to obtain shunt intermediates *in vivo*. Dephosphorylated **8-11** and **8-13** were added to the growth media when THI5p was overexpressed in BL21(DE3) Δ*pdxJ* which cannot biosynthesize PLP. Dephosphorylated **8-13** was the negative test based on the *in vitro* test. Dephosphorylated **8-11** was assumed as the positive test to get the shunt intermediate. However, dephosphorylated **8-**

11 seems to be reduced to PLP *in vivo*. Therefore, the purified THI5p was similar to the inactive THI5p from the *in vitro* assays. This would require dephosphorylated **8-15** to test because the *in vitro* assays showed that it might form shunt intermediates. The advantage to this approach is that that dephosphorylated PLP analogues can be used and the phosphorylation step during the synthesis can be skipped. The experimental analysis is also simpler because there is no co-purified PLP on THI5p. Another possible experiment is to use histidine analogues. Therefore, the possible approach is to use the pyrrolysyl-tRNA synthetase/$tRNA^{Pyl}$ pair to incorporate unnatural histidine at His66 [39].

Lastly, the investigations of the THI5p secondary function could be conducted. Preliminary studies could start by determining the ratio of thiamin and THI5p. Furthermore, phenotype of *S. cerevisiae* Δ*THI5* could be observed with thiamin supplementation. Moreover, it might be possible that THI5p has interacting proteins *in vivo* as the assays showed that THI5p used about 10 equivalents of O_2 to generate HMP-P and caused protein oxidation. In *in vivo*, other proteins might interact with THI5p to control the reaction efficiently, making finding these proteins an attractive avenue of research.

REFERENCES

(1) Settembre, E.; Begley, T. P.; Ealick, S. E. *Current Opinion in Structural Biology* **2003**, *13*, 739.

(2) Jurgenson, C. T.; Begley, T. P.; Ealick, S. E. *Annual Review of Biochemistry* **2009**, *78*, 569.

(3) Begley, T. P.; Chatterjee, A.; Hanes, J. W.; Hazra, A.; Ealick, S. E. *Current Opinion in Chemical Biology* **2008**, *12*, 118.

(4) Settembre, E. C.; Dorrestein, P. C.; Park, J.-H.; Augustine, A. M.; Begley, T. P.; Ealick, S. E. *Biochemistry* **2003**, *42*, 2971.

(5) Settembre, E. C.; Dorrestein, P. C.; Zhai, H.; Chatterjee, A.; McLafferty, F. W.; Begley, T. P.; Ealick, S. E. *Biochemistry* **2004**, *43*, 11647.

(6) Hazra, A.; Chatterjee, A.; Begley, T. P. *Journal of the American Chemical Society* **2009**, *131*, 3225.

(7) Hazra, A. B.; Han, Y.; Chatterjee, A.; Zhang, Y.; Lai, R.-Y.; Ealick, S. E.; Begley, T. P. *Journal of the American Chemical Society* **2011**, *133*, 9311.

(8) Chatterjee, A.; Hazra, A. B.; Abdelwahed, S.; Hilmey, D. G.; Begley, T. P. *Angewandte Chemie International Edition* **2010**, *49*, 8653.

(9) Wightman, R.; Meacock, P. A. *Microbiology* **2003**, *149*, 1447.

(10) Tazuya, K.; Yamada, K.; Kumaoka, H. *Biochimica et Biophysica Acta* **1989**, *990*, 73.

(11) Zeidler, J.; Ullah, N.; Gupta, R. N.; Pauloski, R. M.; Sayer, B. G.; Spenser, I. D. *Journal of the American Chemical Society* **2002**, *124*, 4542.

(12) Rodríguez-Navarro, S.; Llorente, B.; Rodríguez-Manzaneque, M. T.; Ramne, A.; Uber, G.; Marchesan, D.; Dujon, B.; Herrero, E.; Sunnerhagen, P.; Pérez-Ortín, J. E. *Yeast* **2002**, *19*, 1261.

(13) Tazuya, K.; Morisaki, M.; Yamada, K.; Kumaoka, H. *Biochemistry International* **1988**, *16*, 955.

(14) Chatterjee, A.; Schroeder, F. C.; Jurgenson, C. T.; Ealick, S. E.; Begley, T. P. *Journal of the American Chemical Society* **2008**, *130*, 11394.

(15) Dorrestein, P. C.; Huili; Taylor, S. V.; McLafferty, F. W.; Begley, T. P. *Journal of the American Chemical Society* **2004**, *126*, 3091.

(16) Chatterjee, A.; Abeydeera, N. D.; Bale, S.; Pai, P.-J.; Dorrestein, P. C.; Russell, D. H.; Ealick, S. E.; Begley, T. P. *Nature* **2011**, *478*, 542.

(17) Zeidler, J.; Sayer, B. G.; Spenser, I. D. *Journal of the American Chemical Society* **2003**, *125*, 13094.

(18) Grue-Soerensen, G.; White, R. L.; Spenser, I. D. *Journal of the American Chemical Society* **1986**, *108*, 146.

(19) Tazuya, K.; Azumi, C.; Yamada, K.; Kumaoka, H. *Biochemistry Molecular Biology International* **1995**, *36*, 883.

(20) Lai, R.-Y.; Huang, S.; Fenwick, M. K.; Hazra, A.; Zhang, Y.; Rajashankar, K.; Philmus, B.; Kinsland, C.; Sanders, J. M.; Ealick, S. E.; Begley, T. P. *Journal of the American Chemical Society* **2012**, *134*, 9157.

(21) Rexach, J. E.; Rogers, C. J.; Yu, S.-H.; Tao, J.; Sun, Y. E.; Hsieh-Wilson, L. C. *Nature Chemical Biology* **2010**, *6*, 645.

(22) Wasserman, H. H.; Long, Y. O.; Parr, J. *Tetrahedron Letters* **2003**, *44*, 361.

(23) Amblard, M.; Fehrentz, J.-A.; Martinez, J.; Subra, G. *Molecular Biotechnology* **2006**, *33*, 239.

(24) Coin, I.; Beyermann, M.; Bienert, M. *Nature Protocols* **2007**, *2*, 3247.

(25) Saraç, A. S. *International Journal of Chemical Kinetics* **1985**, *17*, 1333.

(26) Vlessis, A. A.; Bartos, D.; Trunkey, D. *Biochemical and Biophysical Research Communications* **1990**, *170*, 1281.

(27) Lee, J.-W.; Helmann, J. D. *Nature* **2006**, *440*, 363.

(28) Traore, D. A. K.; Ghazouani, A. E.; Jacquamet, L.; Borel, F.; Ferrer, J.-L.; Lascoux, D.; Ravanat, J.-L.; Jaquinod, M.; Blondin, G.; Caux-Thang, C.; Duarte, V.; Latour, J.-M. *Nature Chemical Biology* **2009**, *5*, 53.

(29) Lam, A. K. Y.; Hutton, C. A.; O'Hair, R. A. J. *Rapid Communications in Mass Spectrometry* **2010**, *24*, 2591.

(30) Dunn, J. D.; Reid, G. E.; Bruening, M. L. *Mass Spectrometry Reviews* **2010**, *29*, 29.

(31) Tsai, C.-F.; Wang, Y.-T.; Chen, Y.-R.; Lai, C.-Y.; Lin, P.-Y.; Pan, K.-T.; Chen, J.-Y.; Khoo, K.-H.; Chen, Y.-J. *Journal of Proteome Research* **2008**, *7*, 4058.

(32) Ryle, M. J.; Hausinger, R. P. *Current Opinion in Chemical Biology* **2002**, *6*, 193.

(33) Wallar, B. J.; Lipscomb, J. D. *Chemical Reviews* **1996**, *96*, 2625.

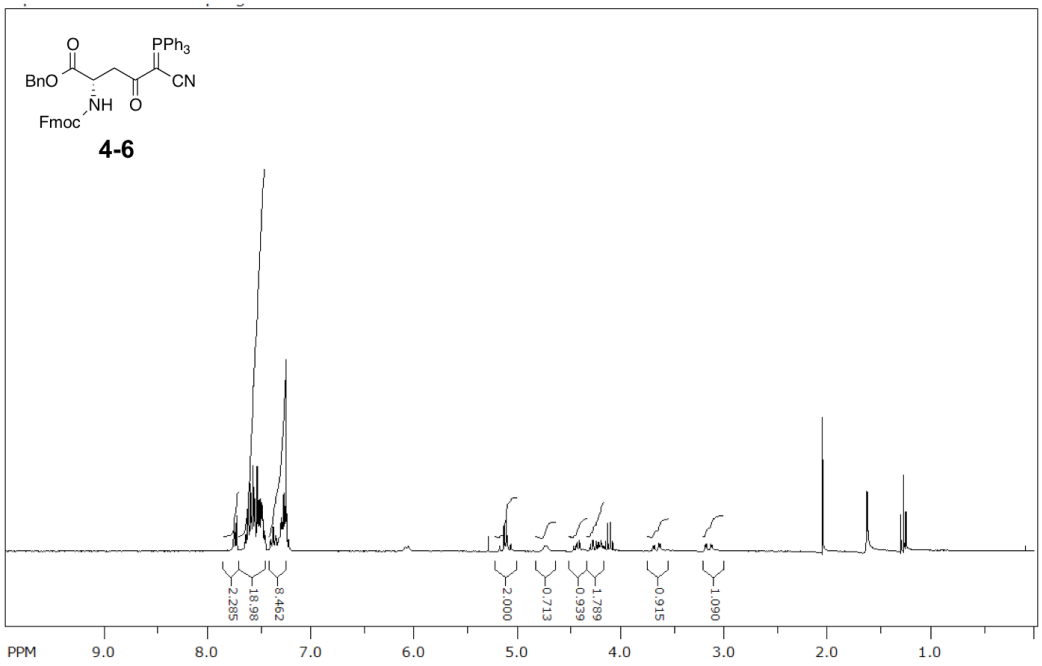
(34) Merkx, M.; Kopp, D. A.; Sazinsky, M. H.; Blazyk, J. L.; Müller, J.; Lippard, S. J. *Angewandte Chemie International Edition* **2001**, *40*, 2782.

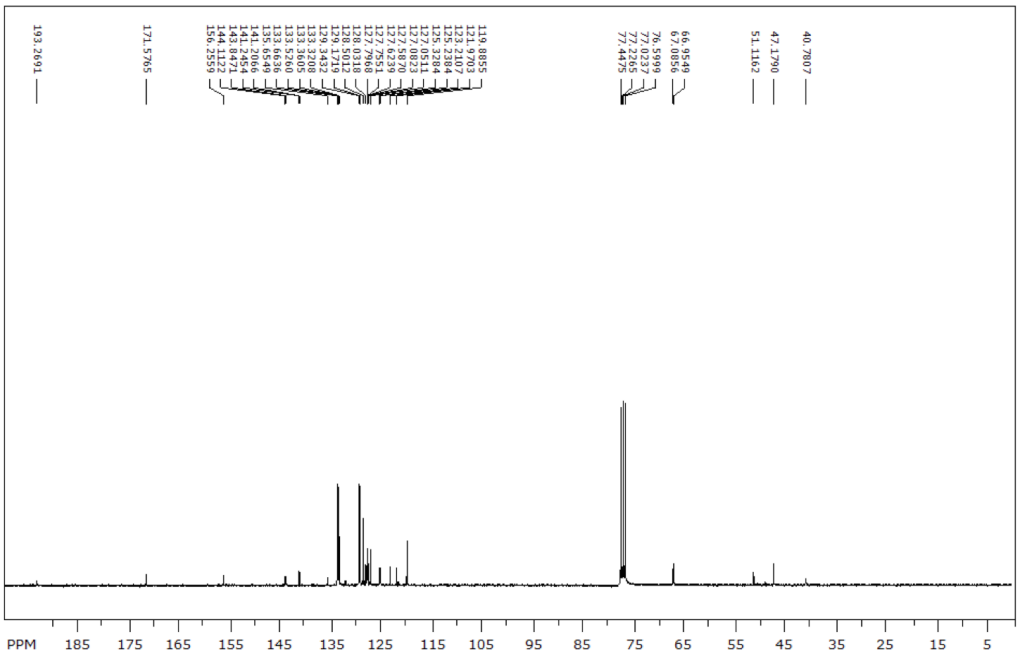
(35) Stubbe, J. *Current Opinion in Chemical Biology* **2003**, *7*, 183.

(36) Dickinson, B. C.; Chang, C. J. *Nature Chemical Biology* **2011**, *7*, 504.

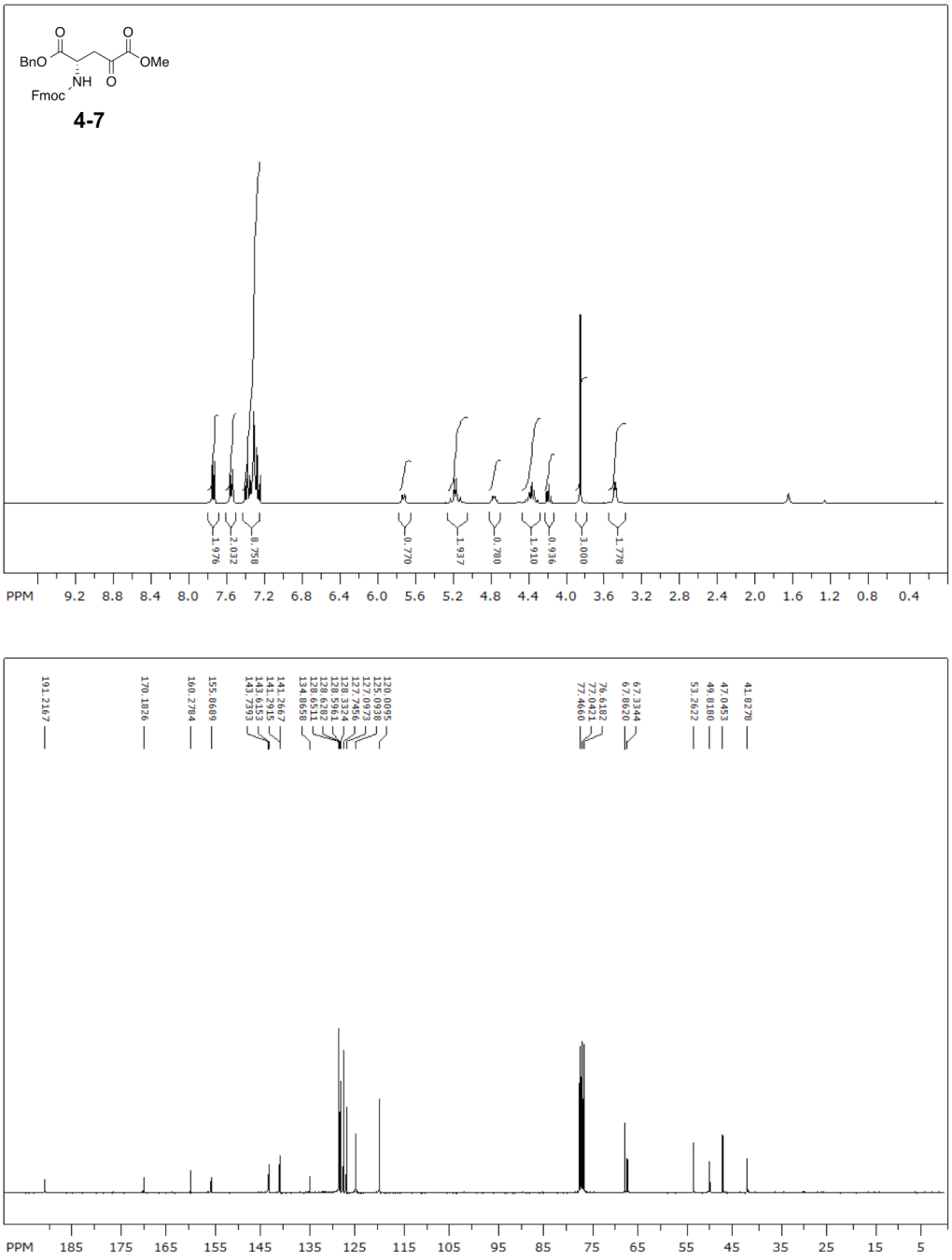
(37) Krężel, A.; Latajka, R.; Bujacz, G. D.; Bal, W. *Inorganic Chemistry* **2003**, *42*, 1994.

(38) Makris, T. M.; Chakrabarti, M.; Münck, E.; Lipscomb, J. D. *Proceedings of the National Academy of Sciences* **2010**, *107*, 15391.

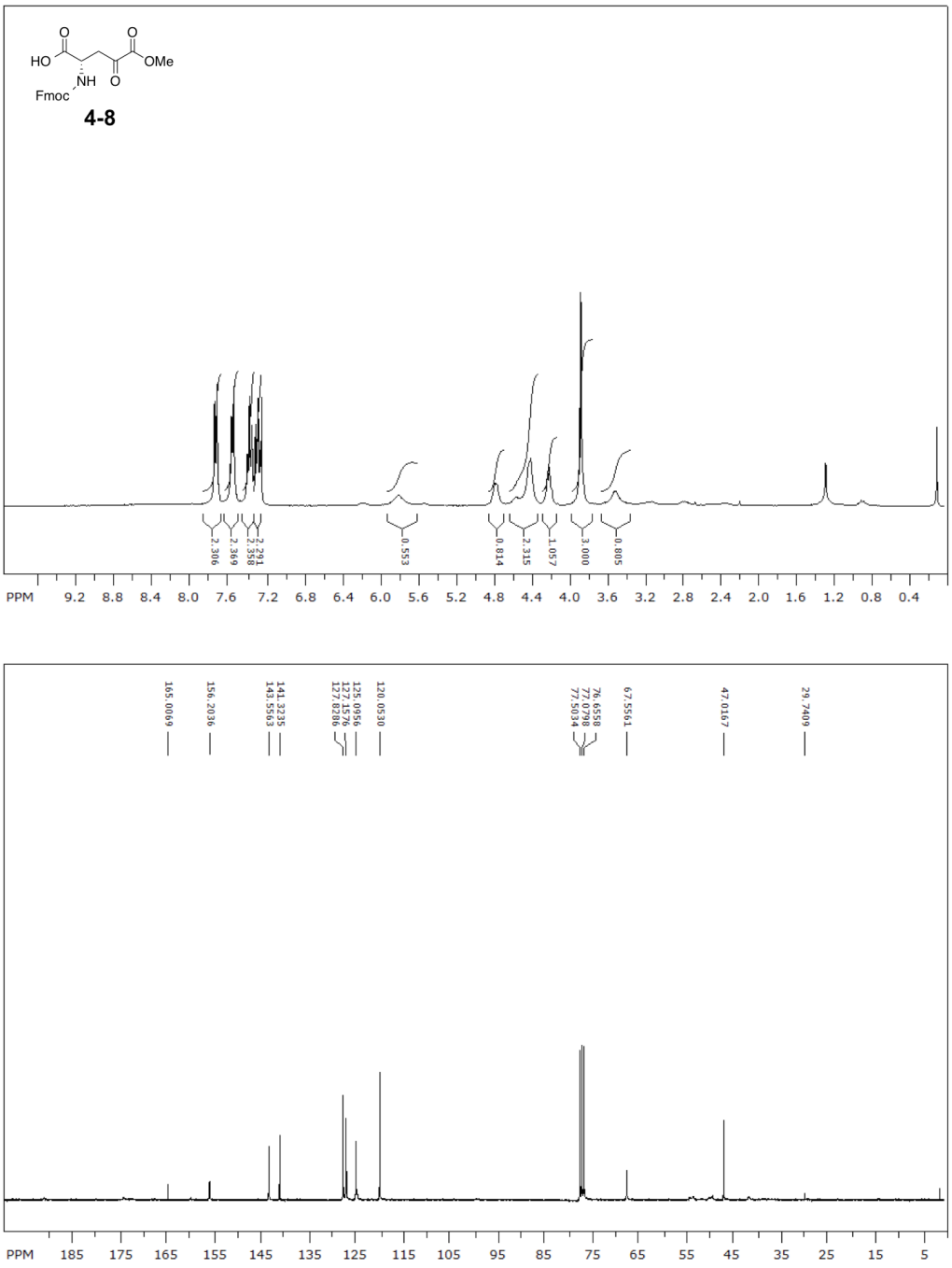




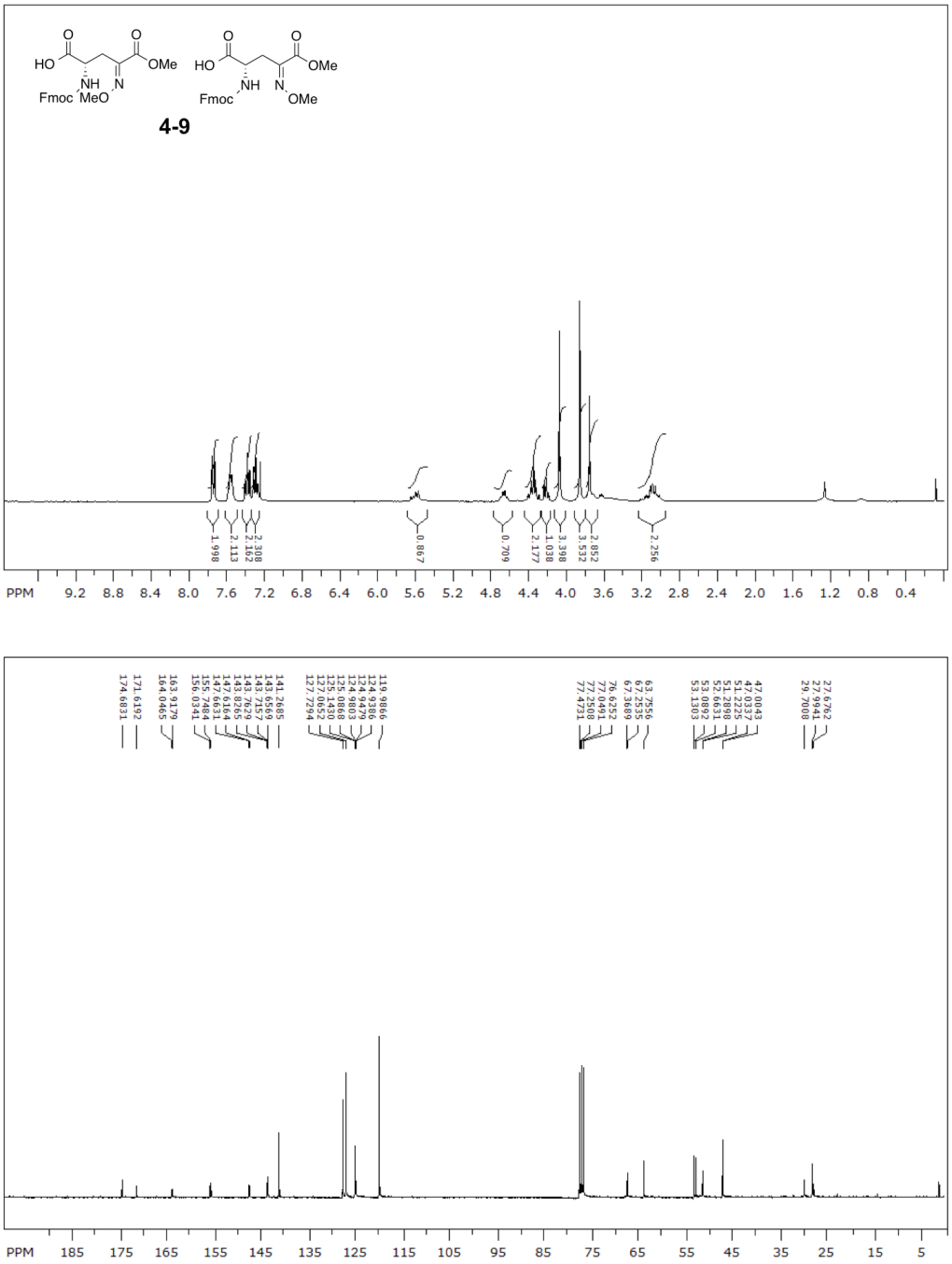
^{1}H- (Top) and ^{13}C- (Bottom) NMR of **4-6**



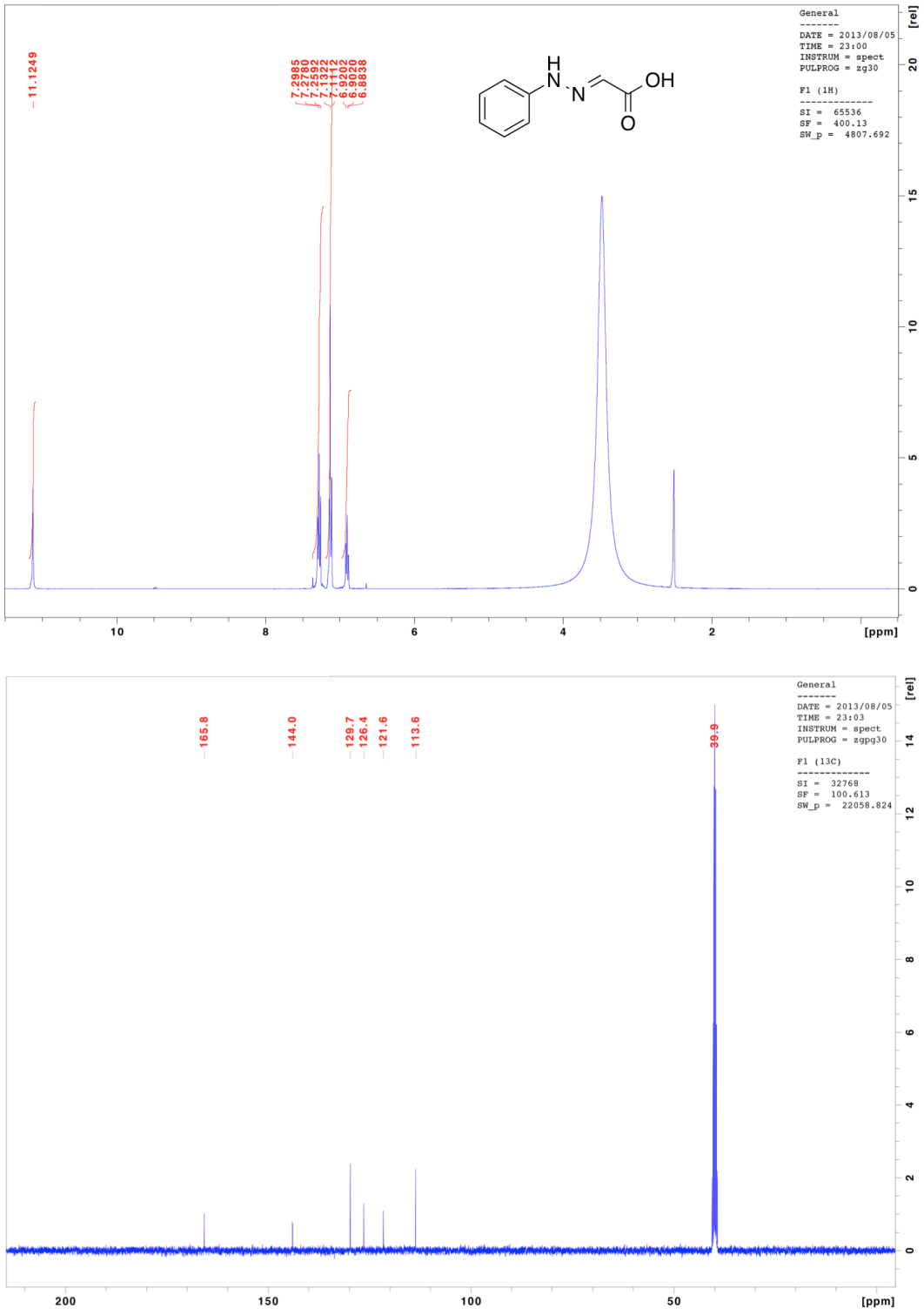
^{1}H- (Top) and ^{13}C- (Bottom) NMR of **4-7**



^{1}H- (Top) and ^{13}C- (Bottom) NMR of **4-8**



^{1}H- (Top) and ^{13}C- (Bottom) NMR of **4-9**



^{1}H- (Top) and ^{13}C- (Bottom) NMR of glyoxylate derivatized by phenylhydrazine